LONGITUDINAL TRENDS IN READING ACHIEVEMENT BETWEEN CHILDREN WITH READING DISABILITY (RD) AND CHILDREN WITH LOW READING ACHIEVEMENT (LRA)

By

Danielle Wexler

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

School Psychology – Doctor of Philosophy

2018

ABSTRACT

LONGITUDINAL TRENDS IN READING ACHIEVEMENT BETWEEN CHILDREN WITH READING DISABILITY (RD) AND CHILDREN WITH LOW READING ACHIEVEMENT (LRA)

By

Danielle Wexler

In elementary school, many children struggle in learning how to read. Some of these struggling readers will be identified to receive special education services as a student with a reading disability (RD), while other students will not be identified to receive such services but will continue to have low reading achievement (LRA). Limited research, however, has examined the differences in the longitudinal reading achievement trajectories between these two groups. Understanding the differences and similarities between children with RD and children with LRA is important for thinking about special education identification processes as well as prevention and intervention efforts. Therefore, the purpose of the current study was to compare the longitudinal reading trajectories of children with RD and children with LRA to determine whether the two groups had different reading achievement in the fall of kindergarten, different patterns of reading achievement from kindergarten through eighth grade, and different reading achievement outcomes in the spring of eighth grade. Primary analyses employed multiple group Latent Growth Modeling (LGM) to examine differences in the reading achievement trajectories between these two groups from kindergarten through eighth grade. Data were drawn from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999 (ECLS-K), which sampled approximately 21,410 children who entered kindergarten in 1998 and followed them through eighth grade. For children identified with RD or LRA between first and fifth grade (n = 1,080), LGM analyses revealed no significant differences between the groups in the fall of kindergarten, significantly different patterns of growth over time, and significant differences in their reading achievement in the spring of eighth grade, with the RD group outperforming the LRA group, even after controlling for gender, socioeconomic status (SES), race/ethnicity, and age. Findings highlight the importance of intervention and special education services to support the reading development of children who face challenges in learning to read. Additionally, findings point to the need for further research to identify why children with RD and LRA had different patterns of reading achievement over time, what the specific reading skill deficits of the students in each of these groups were, and how instruction and intervention can support students' reading development throughout their education.

ACKNOWLEDGEMENTS

I would like to thank several individuals who have contributed immensely to the development of both my dissertation and skills as a researcher, without whom this work would not have been possible. First, I am so very grateful to my advisor and dissertation chair, Dr. Jodene Fine for her continual guidance and encouragement throughout not only the completion of my dissertation but also the entirety of my graduate school training. Thank you so very much for your unwavering support, profound knowledge and wisdom, and constant belief in me. Her mentorship over the past five years has been instrumental in my growth as a scholar, practitioner, and professional in the field of psychology. I would also like to express my deepest appreciation for Dr. Ryan Bowles and his unremitting patience as I worked to conceptualize, carry out, and interpret the statistical analyses for my dissertation. His availability, knowledge, and advice were invaluable in the completion of my dissertation, and I am truly grateful for all that I have learned from him. Next, I would like to thank Dr. Troy Mariage for his willingness to serve on my dissertation committee. His patience, generosity, and expertise have undoubtedly pushed my thinking and knowledge forward, which has helped me to become a stronger researcher. I would also like to thank Dr. Evelyn Oka for her willingness to serve on my dissertation committee. I appreciate her kindness, support, and guidance in sharing her time and wisdom with me.

Additionally, I would like to thank my parents and sister for always believing in me, even when my own self-doubt was at its highest. Thank you for your endless love and support, and without it, I would not be where I am today. Lastly, I would like to thank Michael. There are not enough words to express my gratitude and love for him, and I will forever appreciate your tireless support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Research consistently suggests that being able to read is critical in becoming a successful member of society. Still, learning to read is an exceptionally complex process, and approximately 20-30% of children struggle even after a year of reading instruction (Shaywitz, 2003). Some children who struggle with learning to read will be diagnosed with a reading disability (RD), which is neurological language-based disability that impairs individuals' abilities to retrieve words, articulate words, spell words, and remember facts (Shaywitz, 2003). In the United States, the estimated prevalence of RD for children is 5-17% (Shaywitz & Shaywitz, 2005), with most estimates around 7% (Peterson & Pennington, 2010). Other children who struggle with learning how to read will not receive a RD diagnosis, but are known as students with low reading achievement (LRA). In 2015, approximately 50% of fourth-graders were not proficient in reading and were performing below grade-level expectations (U.S. Department of Education, 2016).

If there are two groups of students struggling to learn to read, then question the becomes, are these two groups inherently different? Researchers have investigated whether there are differences between children with RD and children with LRA in terms of their cognitive and academic functioning as well as their response to instruction/intervention. Regarding differences in cognitive and academic functioning, there are contradictory findings. In general, however, research supports the idea that are minimal to no differences in cognitive and academic functioning (Hoskyn & Swanson, 2000; Stuebing et al., 2002; Ysseldyke, Algozzine, Shinn, and McGue, 1982). In other words, children with RD and LRA all demonstrate a core deficit in phonological awareness, which is believed to be the underlying cause of RD (Fletcher et al., 1994, Stanovich and Siegel, 1994). Although there is evidence to support that students with RD cannot be differentiated from students with LRA measures (Shaywitz, Fletcher, Holahan, & Shaywitz, 1992b; Ysseldyke et al., 1982), there is research to suggest that certain groups of children with reading problems may respond differently to intervention. These studies have found that children who do not respond to intervention are the ones who are in need of the most intensive, individualized services via special education (Denton, Fletcher, Anthony, and Francis, 2006; Lovett et al., 1994; Vaughn et al., 2012).

Given that some children with reading problems may respond more readily to intervention than others, researchers have examined the longitudinal trends in reading development of children with and without RD. Some research has found that the gap between good and poor readers widens over time (Bast & Reistma, 1998; Juel, 1988; McKinney, 1984; McNamara, Scissons, & Gutknecth, 2011), while other research has found that this gap does not widen over time, and may in fact close slightly (Catts, Hogan, & Fey, 2003; Parrila, Aunola, Leskinen, & Nurmi, 2005; Phillips, Norris, Osmond, & Maynard; Shaywitz et al., 1995). Studying children over different developmental periods of time (e.g., kindergarten through third grade versus kindergarten through sixth grade) may have contributed to these conflicting findings. When researchers study longer periods of time, evidence for a widening gap has not been found (e.g., Aarnouste & van Leeuwe, 2000; Phillips et al., 2002). Thus, it is possible that children with RD exhibit a faster rate of progress, which prevents them from falling even further behind (Huang, Moon, & Boren, 2014; Scarborough & Parker, 2003).

Still, other research has found that children with RD progress either at similar rates or more slowly compared to typical readers (Morgan, Farkas, & Wu, 2011; Morgan, Frisco, Farkas, & Hibel; 2010). In other words, these studies suggest that children with RD neither grow at a faster pace nor catch up to their peers without RD. However, in these studies, researchers did not differentiate between children who received special education services for RD and those who were struggling in reading but did not receive special education services. Accordingly, only two studies have examined differences in the reading growth trajectories between children with RD and children with LRA, finding no differences between the groups in reading achievement over time (Francis, Shaywitz, Stuebing, Shaywitz, & Flectcher, 1996; O'Malley, Francis, Foorman, Fletcher, and Swank, 2002).

Consequently, it is important to examine the longitudinal trends in reading achievement between those with RD and those with LRA for two reasons. First, previous research that examined the reading achievement trajectories of these groups concurrently is limited. Thus, little is known about whether the reading achievement of children with RD differs from that of children with LRA over elementary and middle school. The two studies that examined reading growth between the two groups had small samples, used limited observation time periods (e.g., kindergarten through second grade), and failed to consider special education classification. Thus, the current research will allow for a better understanding of whether the reading growth trajectories of children with RD and children with LRA differ from each other over a significant portion of their schooling. Second, understanding the achievement trajectories of those who receive special education services versus those who do not may inform not only future prevention and intervention research but also how the two groups are classified, diagnosed, and labeled within the school setting.

Overall, the purpose of the current study is to test a hypothesized latent growth model to examine the reading trajectories from kindergarten through eighth grade of children who received special education services for RD and non-identified children with LRA between first and fifth grade. Using special education eligibility, this study will seek to evaluate whether children diagnosed with RD are similar or different from children with LRA in terms of their reading achievement at the start of kindergarten, as they grow over time, and at the end of eighth grade. Through the use of a large, representative sample, a long-term longitudinal model, and school-based eligibility, the study will address previous methodological flaws in extant literature, which not only is limited but also fails to consider school-based classification.

CHAPTER 2
LITERATURE REVIEW

Importance of Reading

Research consistently suggests that being able to read is critical in becoming a successful member of society. Reading is important not only for academic success during the school years but also for social and economic advancement later in life (Snow, Burns, & Griffin, 1998). Those with low reading achievement (LRA) often face behavioral, emotional, and social difficulties throughout school (Arnold et al., 2005; Heiervang, Lund, Stevenson, & Hugdahl, 2009; Miles & Stipek, 2006; Morgan, Farkas, Tufis, & Sperling, 2008a), which can lead to more serious antisocial behavior (Svetaz, Ireland, & Blum, 2000). To help prevent and combat the unwanted effects of poor reading achievement, researchers, educational professionals, and policy makers have taken significant steps to implement more empirically-supported reading instruction in schools across the nation.

In the 1990s, reading was pushed to the forefront of the agendas of many government officials and policy makers. These individuals recognized the importance of acquiring reading skills to gain societal success (Miskel & Song, 2004). They also realized that current levels of reading achievement, based on test scores from the National Assessment of Educational Progress (NAEP) were inadequate (Miskel & Song, 2004). Outside the federal government, collaboration between policy makers and researchers allowed for the publication of many popular reports on reading and research-based recommendations for reading instruction, including ones written by the National Research Council (Snow et al., 1998), the Learning First Alliance (1998), the American Federation of Teachers (1998), and the National Reading Panel (2000).

In 2001, President Bush recognized the importance of reading development by including major initiatives in the No Child Left Behind Act (NCLB; 2002). These initiatives were called

Reading First and Early Reading First. As part of the initiatives, funds were allocated to states and school districts in order to prevent reading deficits through the establishment of research-based, high-quality reading instruction for children at high risk for reading difficulties in kindergarten through third grade (Katz, Stone, Carlisle, Corey, & Zeng, 2008). The idea was that early prevention efforts via systematic instruction in reading-related skills within the general education setting could prevent later reading failure in high-risk groups (Katz et al., 2008).

Unfortunately, the results of the Reading First initiative were not promising. The initiative did not improve students' reading comprehension test scores in first, second, or third grade (Gamse, Bloom, Kemple, & Jacob, 2008). Although the results were not positive, the initiative helped to lay the foundation for the emphasis of research-based reading instruction to improve the reading skills of children across the United States.

Reading Development

Learning to read is an exceptionally complex process. More specifically, English has approximately 5,000 different possible syllables, and as an alphabetic language, English relies on using symbols to represent individual parts of a spoken syllable (Snow et al., 1998). The symbols, or letters, used to represent the syllables are also meaningless and abstract, until the reader understands that letters represent syllables, or sound units, in spoken English (Snow et al., 1998). The English writing system is a deep orthography, which means that particular letters can spell several different sounds, because spelling, not phonological representations (i.e., how words sounds), are persevered when writing words (Snow et al., 1998). In other words, in the English language, the same letter or combination of letters can take on different sounds depending on the word in which they are used. Even without considering other cognitive processes that are needed to learn how to read (e.g., attention, memory), the English language poses its own complexities in learning how to read.

Chall's Stages of Reading Development

Despite the complexities of the English language, one way to conceptualize reading development over time is through Chall's (1983) stages of reading development. Chall's (1983) model is a developmental theory that incudes reading stages that are qualitatively different from one another and generally progress in a hierarchical fashion. That is, the skills at one stage must be acquired, at least in part, before moving on to the subsequent stage. Chall's (1983) theory includes six stages, which are described below.

Stage 0. First, Stage 0, or the Pre-Reading stage, occurs from birth to around six years of age. In this stage, children become familiar with books and the concept of stories (Chall, 1983). They learn the letters of the alphabet, how to write their own name, use pencil and paper, and retell stories by looking at the pages of books (Chall, 1983). In addition, they begin to understand basic rhymes and different sounds in words (Chall, 1983). Prior to entering kindergarten, children's early preschool experiences as well their home-literacy environment likely influence their development of literacy skills at this stage (Chall, 1983).

Stage 1. Next, in Stage 1, the Decoding stage, children in kindergarten and first grade begin to learn the relationship between letters and sounds as well as between print and spoken words (Chall, 1983). During this stage, children gain *phonological awareness* (PA), or the general ability to distinguish and manipulate sound units within spoken language (Shaywitz et al., 2008a). Shortly after birth, infants are able to distinguish all of the sounds of any human language, and phonological development continues to be fine-tuned into the early school years (Snow et al., 1998). Children first develop an awareness of spoken whole words, followed by syllables, and then at the phoneme level. When children understand that words are made up of phonemes (e.g., */k/, /a/, /t/* → cat), they have developed *phonemic awareness* (Shaywitz, 2003). Phonemes are the smallest units of spoken language, which combine to form syllables, and English contains about 41 phonemes (National Reading Panel, 2000) Research suggests that the development of phonemic awareness requires at least an introduction to reading instruction before children understand that words are made up of sound units smaller than syllables (i.e., phonemes; Goswami, 1993). Once phonemic awareness is acquired, children learn to recognize that the letters on the page represent the sounds they hear. Understanding the link between spoken and written letters is called the *alphabetic principle*, and attaining this principle allows children to begin reading (Shaywitz, 2003).

In Stage 1, decoding, or word reading, is an important reading outcome to consider (Chall, 1983). In order to read both familiar and unfamiliar words, readers must decode words, which involves converting the letters on the page (i.e., graphemes) into their individual sounds (i.e., phonemes) and blending the phonemes to make a recognizable word (National Reading Panel, 2000). When assessing children's decoding skills, they are asked to read aloud lists of both real words and pseudowords (i.e., clup). Therefore, decoding relies both on children's phonological awareness and alphabetic principle.

Of note, phonological skills have been found to be a strong predictor of later reading skills (Mann & Liberman, 1984; Melby-Lervag, Lyster, & Hulme, 2012; Torgesen, Wagner, Rashotte, Burgess, & Hecht, 1997; Wagner, Torgesen, Laughon, Simmons, & Rashotte, 1993). Even further, Swank and Catts (1994) found that scores on measures of phonological awareness tests were moderately correlated with scores on measures of decoding from the beginning to the end of first grade, and early phonological awareness distinguished poor decoders from strong decoders. More recently, in a study of 514 Canadian children, a measure of phonemic awareness and letter understanding accurately identified at-risk students in kindergarten and predicted their first-grade reading achievement (McNamara, Scissons, and Dahleu, 2005). Consequently, those who struggle to develop these foundational phonological skills will also struggle with decoding words and subsequent reading development at the next stage.

Stage 2. Stage 2, the Confirmation, Fluency, Ungluing from Print stage, is when children in the second and third grades begin to read with increasing fluency (Chall, 1983). Fluency is the ability to read text both quickly and accurately (Armbruster et al., 2010). When fluent readers read, they recognize words automatically, without having to sound them out in their heads (Armbruster et al., 2010). Becoming a fluent reader allows children to begin to focus their attention on comprehension, or understanding of the words on the page, rather than having to spend all of their effort decoding the words (Armbuster et al., 2010). Also important in this stage is vocabulary development (Chall, 1983). As children experience more exposure and practice, their vocabularies continue to grow. During early school years, children's vocabularies expand rapidly, although this expansion is highly variable among individual children (Snow et al., 1998). Furthermore, vocabulary plays and important role in both learning to reading and comprehending text, such that children cannot understand the text if they do not know the meaning of most words (National Reading Panel, 2000).

Stage 3. Subsequently, Stage 3, the Reading for Learning the New: A First Step stage emphasizes reading to learn new ideas and knowledge (Chall, 1983). Chall (1983) argues that at Stage 3, this is done most effectively with text that is clear, simple, and has one viewpoint. At this stage, fourth- to seventh graders are expected to comprehend what they read to learn new information about diverse topics (Chall, 1983). In other words, after third grade, students are expected to read well enough to be able to take meaning away from the text they read. Fittingly, in fourth grade, students often have expanded, more complex subject areas in school (e.g., social studies, science), and, therefore, they are expected to read related texts and learn the material (Chall, 1983).

Correspondingly, comprehension is the ultimate goal of learning how to read, such that if children are able to decode the words and read fluently but not understand the meaning of such words, then they are not actually reading. Important to the idea of comprehension is that in order to understand the meaning of text, readers must be actively thinking and using problem-solving processes while reading (National Reading Panel, 2000). Thus, comprehension is a critical component in becoming a proficient reader, yet it requires readers to decode words and read them fluently before comprehension is possible (National Reading Panel, 2000).

Stage 4. Although comprehension is the ultimate goal of reading development, Chall's (1983) stage theory does not end there. In the next stage, Stage 4, the Multiple Viewpoints stage, middle-schoolers and high-schoolers read more complex material and are asked to apply different viewpoints (Chall, 1983). Students read both expository (e.g., textbook) and narrative texts (e.g., fiction novel), remember what they read, critique it, and apply outside knowledge (Chall, 1983).

Stage 5. The final stage of reading development, Stage 5, is the Construction and Reconstruction – A World View stage (Chall, 1983). In this stage, older adolescents and adults use reading for their own needs both professionally and personally (Chall, 1983). When individuals at this stage read, they construct knowledge for themselves using analysis, synthesis, and judgment (Chall, 1983). With continued maturation, adults will often use reading to fulfill differing needs (Chall, 1983).

Bioecological Model

In addition to the six stages, Chall (1983) recognizes the importance of not only within-child factors but also environmental factors for reading development. She writes, "The scheme [model] focuses on what goes on in the person and in the environment to bring about reading development in each of the successive stages" (Chall, 1983, p. 10). She goes on to highlight the importance of individual within child-factors, the environment (e.g., home, school, community, culture), and the interaction between the two. Accordingly, along with Chall's (1983) stage theory of reading development, one must consider a bioecological model (Bronfenbrenner & Ceci, 1994; Bronfenbrenner & Morris, 1998; Bronfenbrenner & Morris, 2006).

The bioecological model posits that development occurs through complex reciprocal interactions between the child and his/her environment (Bronfenbrenner & Ceci, 1994; Bronfenbrenner & Morris, 1998; Bronfenbrenner & Morris, 2006). There are main four components: process, person, context, and time (Bronfenbrenner & Morris, 1998). First, process includes both proximal processes and distal processes. Proximal processes refer to the interactions between the child and his/her immediate environment, which are responsible for the child's general well-being and competencies (Bronfenbrenner & Morris, 1998). Distal processes refer to interactions between the child's immediate environment and other environments (e.g., child's family access to community resources), but such interactions may only indirectly influence the child. Second, person refers to the individual characteristics of the child, such as age, sex, temperament, and disability (Bronfenbrenner & Morris, 1998). These characteristics can either directly or indirectly influence proximal processes. Third, contexts are the environments with which the child interacts, and these environments influence proximal processes (i.e., interactions between child and environment; Bronfenbrenner & Morris, 1998). As part of context, Bronfenbrenner proposed a nested system that includes different environments that may directly or indirectly influence the child (i.e., microsystem, mesosystem, exosystem, and macrosystem; (Bronfenbrenner & Ceci, 1994). Lastly, time reflects the way in which development changes over time as children age as well as how events affect development differently at different points in time (Bronfenbrenner & Morris, 1998).

Overall, the bioecological model suggests that development occurs through the child's interactions with his/her environment and is influenced by the child's characteristics, surrounding environments, and time. Consequently, consideration of within-child factors and environmental factors is important when examining reading achievement over time. Thus, in the current study, one must recognize that although reading achievement is the main variable of interest, such growth does not occur in a vacuum separate from the child's environment. Some of these environmental factors will be discussed later in the literature review. Nonetheless, in line with Chall's (1983) stages of reading development, the National Reading Panel (2000) has identified the key components of evidence-based reading instruction that are crucial in helping students to achieve their maximum potential in reading.

The "Big Five" Areas of Reading

Corresponding to Chall's (1983) stages of reading development, the National Reading Panel (2000) evaluated existing research to determine which areas of reading instruction should be targeted. To evaluate the research, the National Reading Panel (2000) conducted a meta-analysis and calculated the effects of instruction in specific reading skills (i.e., small effect = 0.20, moderate effect = 0.50, large effect = 0.80). These five areas are often termed the "big five" and include phonemic awareness, phonics, fluency, vocabulary, and comprehension. The results of the meta-analysis are described below.

Phonemic Awareness

According to the National Reading Panel (2000), phonemic awareness is the ability to attend to and manipulate phonemes in spoken words. Results of 52 studies revealed that phonemic awareness instruction was significantly better than other forms of instruction in helping children to acquire phonemic awareness and apply such skills to their reading and spelling (National Reading Panel, 2000). Accordingly, researchers found a large overall effect size of 0.86, with moderate effects on reading outcomes (d = 0.53) and spelling outcomes (d = 0.59; National Reading Panel, 2000). Researchers concluded that when phonemic awareness is taught effectively, such instruction helps children learn to read (National Reading Panel, 2000).

Phonics

Phonics builds upon phonemic awareness by teaching students to apply their understanding of words and their sounds to written text (National Reading Panel, 2000). The ultimate goal of phonics instruction is to teach students to use the alphabetic principle (i.e., letter-sound principle) to decode words and understand what they read (National Reading Panel, 2000). Results of 38 studies indicated that phonics instruction produced moderate effects overall (d = 0.44; National Reading Panel, 2000). Phonics instruction was most effective in improving the decoding skills for regularly spelled words (d = 0.67) and pseudowords (d = 0.60). Phonics instruction also significantly improved oral reading fluency with small effects (d = 0.25; National Reading Panel, 2000). For comprehension, phonics instruction produced significantly greater improvements for kindergarten and first grade students (d = 0.51), but for older students, phonics instruction was less effective (d = 0.12; National Reading Panel, 2000). However, for older students with RD, phonics instruction was effective in improving comprehension (d = 0.32; National Reading Panel, 2000). Phonics instruction was also found to improve spelling for kindergarten and first graders (d= 0.67), but not for children older than first grade (d = 0.09; National Reading Panel, 2000). Although the results support phonics instruction as being effective for helping students learn how to decode words, read fluently, comprehend text, and spell, the effects were greater for children in younger elementary school grades than in later elementary school grades. This further supports the idea that phonics instruction is especially important during early elementary school grades.

Fluency

The National Reading Panel (2000) suggest that reading fluency instruction, which specifically focuses on teaching students to read text with speed, accuracy, and appropriate expression is often not of focus in literacy instruction. Accordingly, there is less of a consensus on how to teach reading fluency. While some believe that independent practice leads to improvement, others believe that guided fluency instruction or intervention is more effective. Nonetheless, the National Reading Panel (2000) found that guided reading fluency instruction had a moderate effect (d = 0.41) on reading achievement for students in kindergarten through fourth grade. More specifically, reading fluency interventions had the greatest effects on accuracy (d = 0.55), followed by fluency (d = 0.44), and then followed by comprehension (d = 0.35; National Reading Panel, 2000). The National Reading Panel (2000) did not find evidence that encouraging children to read on their own would lead to improvements in reading. Thus, this suggests that guided oral reading is most effective for improving children's reading skills.

Vocabulary

When children learn to read, their use of letter-sound correspondence to sound out words will lead to meaningful comprehension only if the word is in their oral vocabulary (National Reading Panel, 2000). If the word is not in their vocabulary, then it will not be understood when encountered in text. As a result, vocabulary instruction is an important component of reading development. Unfortunately, no studies were found that met the National Reading Panel's (2000) criteria for inclusion in their meta-analysis. Nonetheless, informal analysis of studies that did not meet criteria suggests that both direct and indirect vocabulary instruction lead to improvements in reading achievement (National Reading Panel, 2000).

Comprehension

Although students typically acquire strategies to comprehend text informally, explicit, formal instruction in comprehension strategies helps students to become aware of how well they understand what they read (National Reading Panel, 2000). Based on 203 studies included in the meta-analysis, the National Reading Panel (2000) identified seven strategies that are most effective in improving comprehension: comprehension monitoring, cooperative learning, graphic and semantic organizers including story maps, question answering, question generation, summarization, and multiple strategy instruction. The researchers concluded that when teachers demonstrate, explain, model, and implement instruction related to text comprehension, there will likely be improvement in students' ability to understand what they read (National Reading Panel, 2000). Overall, these "big five" reading skills as well as Chall's (1983) stages of reading development help explain how children learn how to read, and they are also supported by a neurological model of reading development, called the dual route model.

The Dual Route Model

Researchers have proposed different theories in order to understand the process of reading development from a neurological perspective. While there are numerous models of reading development, three widely researched models are the parallel-distributed processing model (Seidenberg & McClelland, 1989), the automaticity model (LaBerge & Samuels, 1974), and the dual route model (Coltheart et al., 1993). One of the most widely accepted models of reading is the dual route model, which has been used to explain the process of reading in both children with and without RD (Coltheart et al., 1993).

In this explanatory model, there are two neural pathways that ultimately lead to reading (See Figure 1). Children begin reading using one pathway and progress to a more efficient one, as they become skilled readers (Coltheart et al. 1993). The non-lexical, or phonological route is slower and more labored, requiring children to convert graphemes (i.e., written letters) to phonemes (i.e., individual sound units) in order to sound out words (Coltheart, 2005). For example, the word *cat* must be broken down into its individual phonemes (i.e., /k/ /a/ /t/) in order for children to read, pronounce, and match the word with its meaning. This slower, effortful process is employed when young children are first learning to read, and even as they become more advanced, this route may be utilized when either irregular unfamiliar words and non-words (i.e., *phliked*) are encountered (Coltheart, 2005). As a result, after reading words numerous times through the phonological route, an association between the visual representation of words and their meanings move into children's long-term memories (Shaywitz, 2003).

Consequently, when children become more advanced readers, they use the lexical, or orthographic route. In this orthographic system, children recognize words by their visual patterns and match them with the meanings they have stored in their mental lexicon, or mental dictionary (Coltheart, 2005). Because this system relies on experience with words, it takes time to develop (Shaywitz, 2003). Therefore, children begin to read by using the phonological route, advancing to the orthographic route as they become more skilled readers.

Neuroimaging Research to Support the Dual Route Model

In accordance with the dual route model, neuroimaging research has identified different neural pathways that are involved in reading (See Figure 2). It is important to understand, however, that there is no specific "reading area" in the brain, but numerous studies have identified different regions of the brain that are activated when individuals read (Shaywitz & Shaywitz, 2008b). Three brain pathways are used for different types of readers, where two are used by beginning readers (i.e., left parieto-temporal region and Broca's area) and another that is used by more advanced, fluent readers (i.e., left occipito-temporal region; Shaywitz & Shaywitz, 2008b). First, beginning readers activate the left parieto-temporal region, which has been named the *word analysis* region (Shaywitz & Shaywitz, 2008b). Here, early readers read words by associating letters with their phonemes, or sounds. Second, a region called Broca's area in the frontal lobe of the brain, or more specifically the inferior frontal gyrus, plays a role in *word analysis* and converting written letters to spoken words (Shaywitz, 2003).

Third, the neural pathway involved in advanced reading is the left occipito-temporal region, which has been named the *word form* region, and it allows readers to recognize whole words by their visual patterns and then retrieve them quickly (Shaywitz & Shaywitz, 2008b). With exposure to words, children become more automatic word processors, such that words that were previously processed by the left parieto-temporal region will be processed by the left occipito-temporal region (Shaywitz & Shaywitz, 2008b). Broca's area, therefore, does not play as big of a role for advanced readers' automatic and fluent processing of words (Shaywitz, 2003). Accordingly, as children become more efficient and skilled readers, this region of the brain also becomes quicker at recognizing word forms (Shaywitz et al., 2008a). Although there is a very clear path of typical reading development, which is supported by neuroimaging research, many children often face challenges in the process of learning to read.

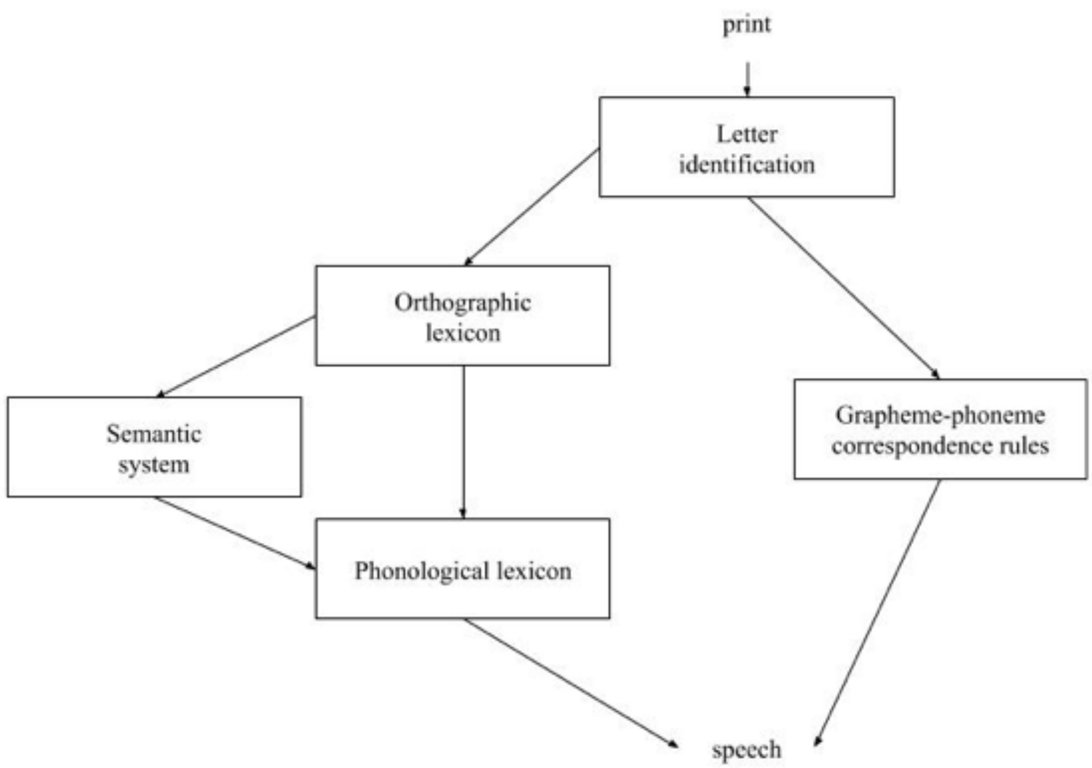

Figure 1. *Dual Route Model* (Coltheart, 2006).

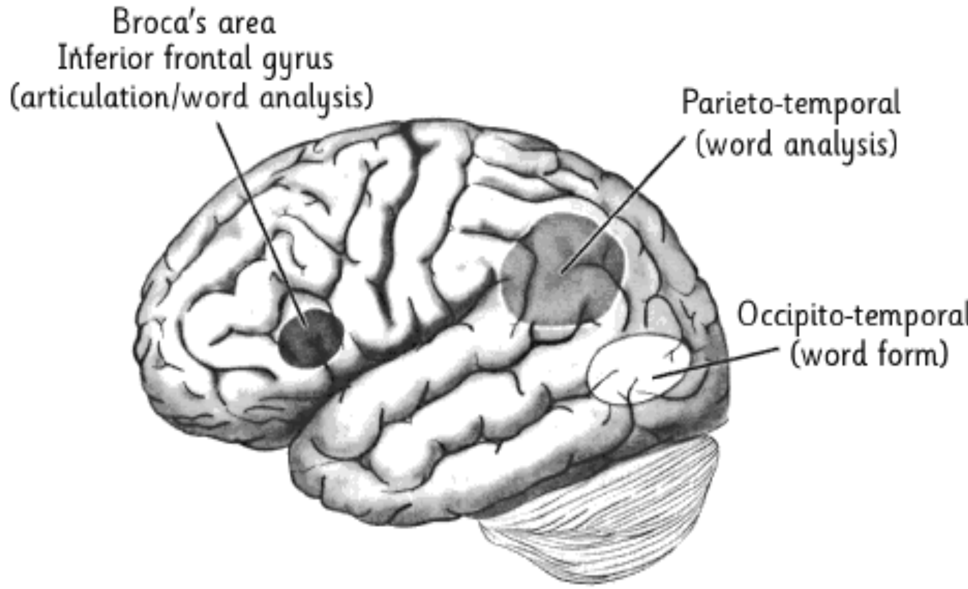

Figure 2. Brain Regions Involved in Reading (Shaywitz, 2003).

Reading Disability (RD)

Reading disability (RD) is a type of learning disability in which students have difficulty making progress in learning to read (Peterson & Pennington, 2010). More specifically, RD is a neurological language-based disability that impairs individuals' abilities to retrieve words, articulate words, spell words, and remember facts (Shaywitz, 2003). RD does not include individuals whose reading difficulties arise from environmental deprivation, sensory impairments, or more severe developmental disorders (Peterson & Pennington, 2010). Even though RD is neurologically based, it is a relatively common disability.

In the United States, the estimated prevalence of RD for children is 5-17% (Shaywitz & Shaywitz, 2005). However, prevalence rates of RD vary depending on the cutoff used to determine whether an individual meets diagnostic criteria. Based on a cutoff for reading achievement at 1.5 standard deviations below the mean, approximately seven percent of the population is identified as having a RD (Peterson & Pennington, 2010). Additionally, since RD is said to encompass at least 80% of all learning disabilities (Shaywitz, 2003), it is estimated that in 2013 about 32% of school-aged children who received special education services were eligible for such services due to RD (U.S. Department of Education, 2015). Moreover, RD has been found to be equally prevalent among boys and girls based on epidemiological studies (e.g., Shaywitz, Shaywitz, Fletcher, and Escobar, 1990); however, in referred samples, boys are more likely than girls to be diagnosed with RD (Rutter et al., 2004).

Etiology of RD

Research suggests a strong neurological component in the development of RD. Neuroimaging research has found neural differences between children with and without RD in support of the dual route model. First, nonimpaired readers at all ages show strong activation patterns in the left parieto-temporal of the brain (i.e., word analysis region) with lesser activation in anterior regions (i.e., Broca's area), suggesting that activation patterns in nonimpaired readers do not change with age (Shaywitz et al., 2002). More specifically, nonimpaired readers (7-18 years) show significantly greater activation in left hemisphere brain regions, such as in the inferior frontal, superior temporal, parieto-temporal, and temporal-middle occipital gyri than do children with RD (Shaywitz et al., 2002). Contrastingly, readers with RD show decreased activation in left posterior regions (i.e., parieto-temporal and occipito-temporal) of the brain (Shaywitz et al., 2002), suggesting that both the word analysis and word form regions are impaired in children with RD. Notably, as result of the underactivation in the left parieto-temporal region, there is underactivation in the occipito-tempotal word form region, because the storage of known words depends on accurate decoding (Pennington, 2009).

Furthermore, Shaywitz et al. (2002) found a positive correlation between age and activation in frontal regions of the brain for children with RD. Young children with RD, as early as six, have been found to have abnormal activation patterns bilaterally, specifically increased activation in bilateral parieto-temporal regions, as opposed to left-side only activation (Specht et al., 2009). As readers with RD get older, activation increases in bilateral anterior regions (Bach et al., 2010), and during adolescence, such children demonstrate a pattern of overactivation in Broca's area (Shaywitz et al., 2002). Shaywitz (2003) suggests that struggling readers compensate for the disruption in the parieto-temporal region in the back of the brain by activating a similar word analysis region in the front of the brain (i.e., Broca's area).

Overall, this neurological evidence indicates that children with RD fail to develop the brain systems involved in phonological processing (i.e., left parieto-temporal region) in the same way as do nonimapired readers. This then prevents brain regions involved in rapid orthographic processing of unfamiliar words (i.e., occipito-temporal region) from developing. Because these brain regions are underactiavted, children with RD try to compensate via bilateral activation of posterior brain regions as well as anterior regions of the brain that are less efficient. Accordingly, this neurological evidence supports the phonological theory of RD, which is discussed in more detail below.

Along with the neurological evidence for RD, it is hypothesized that RD is heritable. Research has found that in large samples of children with RD, over half of the group deficits were due to heritable influences (Peterson & Pennington, 2010). Researchers estimate that 23-65% of children with RD have a parent with RD (Scarborough, 1990), and approximately 40% of their siblings have RD (Pennington & Gilger, 1996).

To further support the heritability of RD, researchers have identified genes on seven chromosomes that play a role in the heritability of RD. The seven genes include: 1p36-p34, 2p16-p15, 3p12-q13, 6p22.2, 15q21, 18p11.2, and Xq27.3 (Peterson & Pennington, 2010). Researchers suggest that RD is caused by the interplay of multiple genes, not just one of them (Peterson & Pennington, 2010). Using behavioral genetic analysis, researchers have also found genetic correlations between component skills of reading. For example, significant genetic correlations were found between word reading and phonological coding as well as word reading and phonemic awareness (Peterson & Pennington, 2010), which means that these component skills are also heritable for those with RD (Gayan and Olson, 2001 as cited in Peterson & Pennington, 2010). Although there is strong evidence to suggest the heritability of RD, environmental factors must also be considered when discussing the etiology of RD.

Because RD is not 100% heritable, environmental factors can also influence the development of RD. Two types of environmental risk factors for RD have been of great focus:

home-literacy environment and instruction quality/type. In regard to home-literacy environment, research has found that parent-child reading predicts later literacy skills (see Phillips & Lonigan, 2005). Other research has found that children who became poor readers in elementary school had less exposure to print as young children; however, it is not clear whether individual child factors or parent factors led to the decreased exposure to print (Peterson & Pennington, 2010). With respect to literacy instruction, research indicates that early intervention that focuses on phonemic awareness, decoding, word reading, and reading fluency can positively affect students who are at risk for reading failure (e.g., Vellutino, Scanlon, Small, & Fanuele, 2006). Overall, research suggests that there is a strong genetic component of RD, yet there are environmental factors that can also influence reading development.

Models of RD

Researchers have proposed a variety of theories to understand the deficits involved in RD. Four such theories of RD are the magnocellular theory (Livingstone, Rosen, Drislane, & Galaburda, 1991; Lovegrove, Bowling, Badcocket, & Blackwood, 1980; Stein and Walsh, 1997), the auditory processing theory (Tallal, 1980; Tallal, Miller, & Fitch, 1995), the double-deficit hypothesis (Bowers & Wolf, 1993; Wolf & Bowers, 1999; Wolf, Bowers, & Biddle, 2000), and the phonological theory (Liberman, 1971; Stanovich, 1985, 1988). Even though there is evidence to support each of these theories, the most supported theory of RD is the phonological theory, which suggests that individuals with RD have decoding, fluency, and/or comprehension difficulties that stem from deficits in phonological processing (Liberman & Shankweiler, 1979, 1991; Snowling, 1981; Stanovich, 1988; Torgesen, 1999). Much of the research that supports the phonological theory has focused on children with RD as having difficulty with phonological processing relative to nonimapaired participants on tests that evaluate phonological awareness and letter-sound decoding (e.g., Blachman, 2000; Fletcher et al. 1994; Liberman, Shankweiler, Fischer, & Carter, 1974; Stanovich & Siegal, 1994; Vellutino, Scanlon, & Spearing, 1995; Vellutino et al., 1996; Wagner & Torgesen, 1987; Wagner, Torgesen, & Rashotte, 1994). Thus, there is a general consensus that poor phonological awareness is the most influential cause of reading difficulties (Vellutino, Fletcher, Snowling, & Scanlon, 2004). Despite the idea that phonological deficits are at the core of RD, both the field of education and the field of psychology have developed their own definitions of RD.

Definitions of RD

There are two widely accepted definitions of RD. First, RD is included in the Individuals with Disabilities Education Act (IDEA, 2004) under the Specific Learning Disability (SLD) category. According to the most recent re-authorization of IDEA (2004), a SLD is

> "a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, that may manifest itself in the imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations, including conditions such as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. Such term does not include a learning problem that is primarily the result of visual, hearing, or motor disabilities, of mental retardation, of emotional disturbance, or of environmental, cultural, or economic disadvantage." (p. 118)

Based on this definition, there are eight categories of SLD: oral expression, listening comprehension, written expression, basic reading skills, reading fluency skills, reading comprehension, mathematics calculation, and mathematics problem-solving. Children with RD typically have a SLD in the areas of basic reading skills, reading fluency skills, and/or reading comprehension. A SLD in the area of basic reading skills reflects a child's difficulty with phonological reading skills (i.e., phonemic awareness and/or phonics). A SLD in the area of reading fluency reflects a child's difficulty with reading text accurately and quickly, and a SLD in reading comprehension, reflects a child's difficulty with understanding what he/she reads. A child can be identified as having a SLD in one or more of these areas. It is likely, however, that when identified as having RD related to fluency and/or comprehension, the child also has difficulty with basic reading skills.

The second definition of RD is from the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; APA, 2013). In the DSM-5, RD is defined under the category of SLD, and is considered a neurodevelopmental disorder. Here, a SLD is a

> "persistent difficulty in reading, writing, arithmetic, or mathematical reasoning skills during formal years of schooling. Symptoms may include inaccurate or slow and effortful reading, poor written expression that lacks clarity, difficulties remembering number facts, or inaccurate mathematical reasoning. Current academic skills must be well below the average range of scores in culturally and linguistically appropriate tests of reading, writing, or mathematics. The individual's difficulties must not be better explained by developmental, neurological, sensory (vision or hearing), or motor disorders and must significantly interfere with academic achievement, occupational performance, or activities of daily living." (APA, 2013b, p. 1)

Similar to the IDEA (2004) definition of SLD, a person can have a SLD in a number of areas. More specifically related to reading, the DSM-5 specifies that a person can have a reading impairment in word reading accuracy, reading rate or fluency, and/or reading comprehension. Accordingly, there is considerable overlap in terms of how RD is defined based on IDEA (2004) and the DSM-5 as well as the way in which children with RD are identified.

Identification of RD

Historically, children with RD were identified based on a discrepancy model, in which a student with a severe discrepancy between his/her reading achievement scores and intellectual ability (i.e., IQ) resulted in a RD diagnosis. The IQ-achievement discrepancy model of RD identification was adopted as part of P.L. 94-142, or the Education for All Handicapped Children Act (1975). Following federal adoption, individual states began adopting the IQ-achievement discrepancy criteria to identify students with RD.

Despite widespread adoption, many researchers argue that using an IQ-achievement-discrepancy model is fraught with problems. One argument against the IQ-achievement discrepancy is that IQ is not a strong predictor of reading achievement and may only account for 25% of the variance in elementary school-aged children's reading performance (Aaron, 1997). Additionally, the relationship between reading and IQ is bidirectional, not unidirectional, such that IQ influences reading and reading influences IQ (e.g., vocabulary development; Aaron, 1997). Third, the IQ-achievement discrepancy neither predicts one's response to intervention nor provides guidance with recommendations for educational planning (Aaron, 1997; Gresham & Vellutino, 2010). Next, the IQ-achievement discrepancy model is often thought of as a "wait to fail" model, such that intervention is delayed until the discrepancy between IQ and achievement becomes large enough that the student qualifies for special education services (Fletcher, Coultier, Reschly, & Vaughn, 2004). Finally, and perhaps most importantly, the IQ-achievement discrepancy assumes that there are inherent differences, in terms of cognitive processes, between children with RD and children without RD (Aaron, 1997). This issue about cognitive differences between children with RD and those with LRA who may not meet discrepancy criteria is discussed further below. In sum, researchers who argue against the use of IQ-achievement discrepancies for RD classification contend that the discrepancy paradigm either overidentifies or underidentifies children as having RD (Fletcher et al., 1994; Fletcher et al. 2004; Vellutino, Scanlon, & Lyon, 2003).

To address problems with RD identification resulting from the IQ-achievement discrepancy paradigm, changes have been made to RD identification definitions in both IDEA (2004) and the DSM-5. More specifically, IDEA (2004) states that a severe discrepancy between IQ and achievement is no longer required to diagnose RD, and it is acceptable to use a process that is based on the child's response to scientific, research-based intervention. According to the DSM-5, multiple sources of data, including developmental, medical, educational, and familial history records; previous academic reports; observations from teachers; and standardized assessment scores must be considered along with the child's response to intervention in order to diagnose RD (APA, 2013). Also, the DSM-5 states that academic skills exist along a continuum, and there is no one cut-point that can be used to differentiate children with and without RD (APA, 2013). Given these new and alternative ways of identifying children with RD, the IQ-achievement discrepancy model is being used less frequently.

More recent models of RD identification. To replace the IQ-achievement discrepancy model, RD is now commonly diagnosed based on the progress a student makes towards achieving state age- or grade-level standards after receiving evidence-based intervention (EBI) in reading (Gresham, 2007; Kavale & Flanagan, 2007). This process is known as response to intervention (RTI), and it involves the use of a systematic problem-solving process to identify specific reading deficits, implement EBIs, and monitor growth after intervention implementation (Gresham, 2007; Kavale & Flanagan, 2007). If a student does not make adequate progress after receiving an EBI for a designated amount of time, he/she may be diagnosed with an SLD. The RTI model is based on research that suggests children with the most impaired reading skills are those who are the hardest to remediate, and thus require the most intensive supports (Vellutino et al., 2004). This idea is discussed in more detail in a subsequent section.

Another alternative method in which RD is diagnosed is using a pattern of strengths and weaknesses (PSW). Here, the student's academic performance and/or intellectual ability is considered in light of his/her age, intellectual development, or state grade-level standards (Hale, Kaufman, Naglieri, & Kavale, 2006). If a student exhibits a significant weakness in reading and strengths in other areas, then he/she may be diagnosed with RD (Hale et al., 2006).

Although these alternative methods of RD identification are common practice in schools, there is still a great deal of heterogeneity in the diagnosis of RD in both practice and research. First, it is important to note that individual states have the right to determine which identification model they will use, and, in fact, as of 2013, only 16% of states required the use of RTI for RD identification in their special education laws (Maki, Floyd, & Roberson, 2015). Accordingly, some schools still use an IQ-achievement discrepancy model of identification, while others use a PSW model or a RTI framework. Even within each identification model, states and school districts vary in their operationalization of certain parts of the identification model. For example, for those using an IQ-achievement discrepancy model, some may say that one's achievement has to fall at least two standard deviations below his/her IQ, while others will say the difference only needs to be 1.5 standard deviations. Similarly, in an RTI framework, different criteria will be used to determine what it means to make adequate progress in response to an intervention. For example, the number of weeks for which the intervention must be implemented or the number of interventions that must be tried before making an eligibility determination often varies.

Similarly, in research studies, different criteria are used to diagnose RD. Some studies may adhere to a discrepancy model, while others use achievement cutoff scores (Shaywitz et al., 1992b). Even for those who use cutoff scores there is variation in diagnostic procedures, as some use stricter criteria (e.g., below the 10^{th} percentile), while others use more liberal criteria (e.g., below the 30^{th} percentile). Thus, in both research and practice, there has been considerable variability in the way in which children with RD are identified, influencing not only which children are diagnosed with RD versus LRA but also who receives extra services in school.

Low Reading Achievement (LRA)

Despite these recent changes in RD definitions and identification, there are many children in school with low reading achievement (LRA) who do not receive special education services. In fact, in 2002, only 33% of fourth grade students who took a national standardized achievement test (i.e., NAEP) were considered proficient or advanced in reading (U.S. Department of Education, 2016). In 2015, the percentage of students proficient or advanced in reading increased to 40% (U.S. Department of Education, 2016). Although this was a significant increase, 34% of children performed at the basic reading level and 26% of children performed at the below basic level in 2015 (U.S. Department of Education, 2016). Thus, over half of fourth-graders sampled in 2015 did not meet grade-level expectations in reading.

These low achieving children, however, are not necessarily identified to receive special education services. Accordingly, there are three possible reasons to explain why children with LRA may not be identified to receive special education services as a result of RD. First, these children may have been evaluated for special education services but were determined ineligible. They may have been found ineligible, because they did not meet the school's eligibility determination criteria. For example, a school using an achievement-IQ discrepancy model of identification may fail to find a student eligible if there is not a large enough discrepancy between his/her achievement and IQ. Similarly, a school using a RTI model of identification may fail to find a student eligible if he/she responds to intervention. A second reason to explain why children with LRA may not be identified to receive special education may be because of other exclusionary factors, such as environmental (e.g., lack of adequate instruction) and/or cultural factors (e.g., language) that account for the LRA. Lastly, children with LRA may mistakenly not be identified as result of going unnoticed or unreferred.

These low achieving children who are not identified for special education services are often called garden-variety poor readers, meaning that their reading achievement scores are consistent with their ability (Stanovich, 1986, 1988, 1991). In one of first studies that investigated LRA, Rutter and Yule (1975) used the term reading backwardness to describe children with LRA, in contrast to specific reading retardation to describe children with RD. From their research, the authors proposed that children with general reading backwardness were different from children with specific retardation (Rutter and Yule, 1975). However, research has since discredited some of their findings as a result of methodological flaws, which are described further below (e.g., Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992a; van der Wissel & Zegers, 1985). In consequence to these original studies, there has been a significant amount of work to determine what it means to be a student with LRA and how this may differ from a student diagnosed with RD.

Distribution of Reading Achievement

In examining how LRA may differ from RD, it is important to point out the research that supports a normal distribution of reading achievement (Shaywitz et al., 1992a). Originally, it was suggested that RD was distinct from other, more generic reading problems. More specifically, Rutter and Yule (1975) argued that reading ability was bimodally distributed, or that RD represented the extreme lower tail of the distribution with its own hump. However, Rutter and Yule's (1975) results were complicated by methodological issues. First, the measures of reading achievement likely imposed a ceiling on reading scores, which may have skewed the data and created this lower hump (Shaywitz et al., 1992a). Second, Rutter and Yule (1975) did not apply any exclusionary criteria to their study, such that 36% of the children with LRA had diagnosed or suspected neurological disorders (e.g., cerebral palsy, epilepsy), and many also had IQ scores consistent with intellectual disability. Including children with neurological disorders may have inaccurately represented the group of children with LRA, making them appear more distinct from the RD group than they actually were. Accordingly, the bimodal distribution model of reading achievement has not been replicated by other researchers (e.g., Lewis, Hitch, & Walker, 1994; Jorm, Share, Matthews, & Matthews, 1986; Rodgers, 1983; Stevenson, 1988; Shaywitz et al., 1992a).

Shaywitz et al.'s (1992a) research suggests that reading ability is normally distributed and exists on a continuum. By examining IQ-achievement discrepancies of 414 children, researchers determined that no distinct cutoff point can distinguish children with RD from children with normal reading achievement (Shaywitz et al., 1992a). In other words, children with RD represent the lower end of the normal distribution and children with no reading difficulties represent the other end of the distribution. Those with LRA, who do not have a diagnosis of RD, therefore, fall somewhere between those with severe RD and those with high reading ability. Even further, Shaywitz et al. (1992a) found that only 28% of children classified as RD in first grade remained classified as RD in third grade, suggesting that there is some instability and fluidity in RD diagnosis. Consequently, this research provides evidence that children with RD are not necessarily discrete from children with LRA, and ultimately, Shaywitz et al. (1992a) state that RD "is not an all-or-none phenomenon but, like hypertension and obesity, occurs in varying degrees of severity" (p. 149). In light of these findings, Shaywitz et al. (1992a) recommend that educational professionals reconsider the provision of special education services that is based on an arbitrary cutoff, so that those who do not meet these criteria still receive support.

Neurocognitive Differences Between RD and LRA Profiles

In further support of this idea that reading ability exists on a continuum, researchers have examined differences across groups of children with RD who meet IQ-achievement discrepancy criteria and those with LRA who do not meet discrepancy criteria. Overall, there is conflicting evidence about whether children with RD and low achievers can be differentiated. On the one hand, Fuchs and colleagues (Fuchs et al., 2002) argue that RD and LRA groups can reliably be differentiated using cognitive and achievement variables. In fact, Fuchs et al. (2002) found an average effect size of 0.61 to support their argument that the two groups can be distinguished based on their reading achievement and IQ scores. However, Stuebing et al. (2002) points out that Fuchs et al. (2002) used the same variables to both differentiate the groups and to define the groups, which may have created inaccurately high effect sizes.

On the other hand, researchers have failed to find differences in cognitive abilities and achievement when comparing the two groups. For example, Shaywitz, Fletcher, Holahan, and Shaywitz (1992b) did not find any differences across nine cognitive domains, including phoneme deletion, visual-spatial deletion, verbal short-term memory, nonverbal short-term memory, speech production, vocabulary/word finding, rapid naming, visual-motor, and visual attention between children who met RD discrepancy criteria and those with non-discrepant LRA. Similarly, an early study by Ysseldyke et al. (1982) compared 50 fourth-grade children with RD who received special education services to 49 fourth-grade children with LRA who did not receive special education services on 49 psychometric measures. Measured domains included cognitive (e.g., WISC-R), achievement, perceptual-motor, and behavior/emotion (Ysseldyke et al., 1982). The results indicated that there was an 82-100% overlap between the groups' scores on all of the administered measures with a median overlap of 96% (Ysseldyke et al., 1982). Thus, these authors argue that there were no clinical differences in the performance of these two groups across a wide range of psychometric measures (Ysseldyke et al., 1982). In this study, however, demographic variables were not controlled, and the discrepancy criteria used by the school to make RD diagnoses is unknown. In addition, due to the cross-sectional design of this study, one is not able to determine whether the overlap between these two groups remains constant or whether it changes over time.

Despite the limitations of the Ysseldyke et al.'s (1982) study, two independent meta-analyses have demonstrated that effect sizes on measures of achievement and cognitive functioning between children with RD who meet discrepancy criteria and children with non-discrepant LRA are at best small (Hoskyn & Swanson, 2000; Stuebing et al., 2002). More specifically, based on a meta-analysis of 19 studies, Hosykn and Swanson (2000) found no clinically significant differences (i.e., small effect sizes) on measures of phonological processing (i.e., pseudoword decoding, rapid naming) between the RD and LRA groups, suggesting that a phonological processing deficit is at the core of underlying reading deficits in children both with

IQ-achievement discrepant RD and those with non-discrepant reading problems. Perhaps more importantly, when all of the cognitive measures (i.e., syntactical knowledge, lexical knowledge, visual-motor skills, spatial processing) along with the phonological processing measures were included in the model with control variables (i.e., verbal IQ, age), any significant differences between the two groups on measures of lexical knowledge, syntactical knowledge, and visual-spatial processing disappeared (Hosykn & Swanson, 2000). Thus, the authors suggest that differences in these cognitive processes were an artifact of certain control variables (e.g., age, verbal IQ), so the cognitive processes themselves do not necessarily distinguish the two groups (Hosykn & Swanson, 2000).

In a meta-analysis of 46 studies, Stuebing et al. (2002) found similar results to those of Hoskyn and Swanson (2000). In comparison to Hoskyn and Swanson (2000), Stuebing et al. (2002) used more lenient criteria to define the groups. Even though Stuebing et al. (2002) used less stringent inclusionary criteria, they found similar results to Hoskyn and Swanson (2000). Overall, negligible effects across behavior and achievement domains and small effects for the cognitive ability domain were found (Stuebing et al., 2002). Related to achievement, Stuebing et al. (2002) found no meaningful differences on the four constructs most closely related to reading, including phonological awareness (d = -0.13), rapid naming (d = -0.12), vocabulary/lexical skills (d = -0.10), and verbal short-term memory (d = 0.10; Stuebing et al., 2002). Within the cognitive domain, some differences between the groups (i.e., fine motor, concept formation, spatial, planning, perceptual-motor, nonverbal short-term memory), with small to medium effects, were found (Stuebing et al., 2002). However, Stuebing et al. (2002) argue that (1) none of these constructs have been shown to be strongly linked to reading ability, (2) other research indicated poorer performance on similar measures by poor readers in comparison to good readers irrespective of how poor readers were defined (Doehring, 1978; Lyon, 1985; Rourke, 1975; Vellutino, 1979), and (3) when more complex, multivariate models include verbal and nonverbal measures, the nonverbal measures do not account for much of the variance in reading achievement (Fletcher, Foorman, Shaywitz, & Shaywitz, 1999; Shankweiler & Crain, 1986; Share & Stanovich, 1995). Moreover, the RD and LRA groups performed similarly on measures most closely linked to reading proficiency, including phonological awareness, rapid naming, verbal short-term memory, and vocabulary/lexical skills, which supports the phonological theory of RD, or that problems with phonological awareness underlie reading difficulties (Stuebing et al., 2002).

Two additional studies have found evidence to support children with RD and LRA as having a core deficit in phonological awareness. First, 79-98% of the sample was consistently identified as RD using different criteria of ability, achievement, and severity, making it impossible to accurately separate subgroups of poor readers (Fletcher et al., 1994). In other words, it was impossible to separate different subgroups of poor readers based. Second, Fletcher et al. (1994) suggest that the main correlate of reading difficulties was always related to linguistic processes, which further supports the phonological hypothesis of RD and the idea that reading ability exists on a continuum.

Similarly, in a study of 907 children with either RD or LRA, Stanovich and Siegel (1994) found that across seven measures of phonological processing, only two were significantly different between the two groups, but effect sizes were small (-0.098 and -0.091). Although the effects were small, the children with RD outperformed the LRA group, which is further evidence to support that both children with RD and LRA have impairments in phonological awareness (Stanovich & Siegel, 1994). However, Stanovich and Siegel (1994) found that the RD group outperformed the LRA group on measures of working and short-term memory, which are domains they argue reside outside of phonological processing. Although there are some discrepancies regarding the areas in which children with RD and LRA may differ, the most prevalent finding from these studies and meta-analyses is that both groups exhibit a core phonological processing deficit, and such groups cannot be distinguished based on this deficit. Accordingly, given that RD and LRA groups experience similar reading difficulties, it is necessary to examine the services and supports provided to these students and whether there are differential effects of intervention.

Services and Supports for Students with RD and LRA

By and large, a big part of the debate about differentiating students with RD from students with LRA is related to how they respond to services and intervention. Controversy exists as to whether children with LRA should be provided the same intensive services/intervention as children with diagnosed RD (e.g., Allington & McGill-Franzen, 1989; Shaywitz et al., 1992a; Siegel, 1992). Prior to the reauthorization of IDEA in 2004, children who were found eligible to receive special education services as a student with RD based on a discrepancy model received extra support via special education. This extra support varied from placements in a self-contained classroom, pull-out services in the resource room, or push-in services within the general education classroom.

Also prior to the reauthorization of IDEA in 2004, for those students who did not qualify for special education, they were likely provided with remedial help within their classroom under the guidance of the general education teacher. However, given the demands of a large classroom, it was difficult for teachers to find the time to provide the instruction at its needed intensity (Mathes & Torgesen, 1998). It is also possible that children who were found ineligible for special education would be eligible to receive additional help under the school's Title I program, which provides federal financial support for the education of economically disadvantaged children, usually in pull-out programs (Allington & McGill-Franzen, 1989). Accordingly, there was not only an obvious distinction between the services provided to students with diagnosed RD and those with LRA but also the effectiveness of both types of services may be questionable.

Research has examined the reading support and services provided to students with RD in special education. First, a main concern is that children in special education have lower achievement outcomes than children in general education classrooms. One study that specifically investigated differences in grades between ninth-grade students with RD and those with LRA found that students with RD had significantly lower grades in science, health, and overall GPA (Donahoe & Zigmond, 1990). Second, research has found that there is a significant amount of variability in the time that students in special education actually spend in reading instruction. According to Allington and McGill-Franzen (1989), some students in special education receive nearly no reading instruction at all, while others receive significantly more reading instruction than their peers in general education. In some instances, students in special education who spent time in both the resource room and general education classroom were found to receive the same amount of reading instruction as their peers in the general education, such that these students did not receive additional or supplemental reading instruction (Hayne & Jenkins, 1986). Relatedly, one study found that children with RD who were mainstreamed in general education classrooms received significantly fewer minutes of reading instruction over the school day compared to students who received Title I programming (Allington & McGill-Franzen, 1989).

Furthermore, researchers have found that the quality of reading instruction in special education classrooms is no better or more intensive than the instruction provided in general education classrooms (Dworet, 1987; Morsink, Soar, Soar, & Thomas, 1986; Vaughn, Levy, Coleman & Bos, 2002). One study found that special education for children with RD offered a smaller proportion of active teaching and a larger proportion of seat work activities compared to general education classrooms and Title I programs (Allington & McGill-Franzen, 1989). Still, students in both special education and Title I pull-out programs have been found to receive disjointed and inconsistent programs that do no align with the general education reading curriculum (Johnston, Allington, & Afflerbach, 1985; Moody, Vaughn, Hughes, & Fischer, 2000; Spear-Swerling & Sternberg, 1996; Vaughn, Moody, & Schumm, 1998).

More recently, a study was conducted to examine the effectiveness of reading instruction in resource rooms for students with learning disabilities (Bentum & Aaron, 2003). In the sample, students were determined eligible to receive special education services based on the IQ-achievement discrepancy model (Bentum & Aaron, 2003). There were 230 students in grades one through seven who received at least three years of services, and 164 students in grades one through seven who received at least six years of services (Bentum & Aaron, 2003). Students in the sample spent between 40 and 120 minutes in the resource room each day (Bentum & Aaron, 2003). Although the researchers did not control for any confounding variables, results indicated that those who received instruction in the resource room for three or six years experienced no improvement in reading achievement (decoding and comprehension), but experienced a significant loss in spelling performance (Bentum & Aaron, 2003).

Similar to the concerns raised regarding the effectiveness of special education services, government officials, policy makers, and researchers have also presented more general concerns about the ability of the nation’s schools to teach all children how to read, not just those with diagnosed disabilities. As described at the outset of this literature review, before the implementation of RTI and the movement towards systematic research-based instruction, there were significant concerns regarding the reading achievement of students in the United States. To ameliorate these concerns, Mathes and Torgesen (1998) proposed that in schools, there must be a unique balance between "(a) early identification of children at risk for reading failure followed by (b) carefully designed intense early reading instruction incorporating systematic, explicit instruction in alphabets reading skills, balance with meaningful experiences with authentic texts and writing and (c) continued support beyond the initially acquisition of reading skill to ensure continued growth into the upper grades" (p. 319). From the work of researchers as well as from federal initiatives to improve reading achievement both at the policy level (e.g., NCLB, 2002) and at the implementation science level (e.g., National Reading Panel, 2000; Snow et al., 1998), changes have been made to the way in which reading instruction and intervention is delivered to all students, including those with LRA and RD.

Differential Response to Intervention

When IDEA was reauthorized in 2004 and RTI frameworks for eligibility decisions became more popular, researchers started to investigate the validity of this model and its ability to differentiate students with RD who would require special education services from those with LRA who would respond to evidence-based, intensive intervention in the general education setting. First, however, it is necessary to define what it means to be unresponsive to intervention. Non-responders are those who do not make adequate reading progress despite receiving evidence-based instruction (O'Conner & Klingner, 2010). However, the field has not yet come to a consensus on what defines adequate reading progress (McMaster, Fuchs, Fuchs, & Compton, 2005), and researchers vary on how they define unresponsiveness (i.e., performing within the 10^{th}-50^{th} percentile after intervention; Torgesen, 2000; Vellutino, Scanlon, Zhang, &

Schatschneider, 2008). The argument about unresponsiveness is important, nonetheless, because it lies at the foundation for determining which students should be classified as RD and subsequently receive special education services. In other words, the argument is that children who respond to evidence-based intervention do not need special education services, while those who continue to have difficulties even after receiving intervention should be determined eligible.

Vellutino and colleagues have conducted seminal work in this area. Vellutino et al. (1996) and Vellutino, Scanlon, and Lyon (2000) evaluated whether those with reading difficulties caused by limited early literacy experiences and/or less than adequate literacy instruction responded differently to intervention than those with reading difficulties as a result of underlying cognitive deficits related to reading ability (e.g., phonological awareness). These researchers conducted a longitudinal study in which children's reading achievement and related cognitive abilities were assessed from kindergarten through fourth grade (i.e., before and after they received intervention; Vellutino et al., 1996). After one semester, the children who received the intervention were ranked based on their growth, and four groups were formed: very limited growth (VLG), limited growth (LG), good growth (GG) and very good growth (VGG; Vellutino et al., 1996). The results suggested that the four groups did not differ on measures of intelligence, and none of the groups differed from the group of normal readers with average IQ (Vellutino et al., 1996). Contrastingly, children who were difficult to remediate differed significantly from the easily remediated children on measures of phonological awareness (e.g., knowledge of letter names and sounds, phoneme awareness, letter-sound decoding, verbal memory; Vellutino et al., 1996). Hence, measures of intelligence were unable to distinguish different degrees of struggling readers, but their performance on reading-related measures after intervention did distinguish them as readily remediated or difficult to remediate.

The findings of Vellutino and colleagues' studies in 1996 and 2000 were further supported by more recent research (Vellutino et al., 2006; Vellutino, Scanlon, Zhang, & Schatschneider, 2008). Here, kindergarten children at risk for reading difficulties were randomly assigned to receive either reading intervention or remedial services in school as usual (Vellutino et al., 2006). After receiving the intervention for one year, the students in first grade were grouped based on whether they were still at continued risk (Vellutino et al., 2006). Then, half of those who were at continued risk received additional one-to-one intervention throughout first grade, while the others received school-based intervention as usual (Vellutino et al., 2006).

First in regard to these findings, approximately 84% of the at-risk children became average readers after either the kindergarten intervention or both the kindergarten and first grade intervention combined (Vellutino et al., 2006). Second, 58% of the children who continued to be at risk after kindergarten performed in the average range on all reading outcome measures at the end of first, second, and third grades (Vellutino et al., 2006). Nevertheless, 42% of the children in the continued risk group performed below average on reading outcome measures at the end of second and third grade, even though they performed in the low average to average range on these measures at the end of first grade (Vellutino et al., 2006). These children whose performance decreased over time were called difficult to remediate, while those whose performance became average after the kindergarten and first grade intervention were called less difficult to remediate (Vellutino et al., 2006). In all, only 16% of the original kindergarten at-risk sample demonstrated substantial reading difficulties at the end of second and third grades (Vellutino et al., 2008).

Interestingly, none of the first-grade baseline scores accounted for significant variance in the reading measures (Vellutino et al., 2006). This suggests that the children in the difficult to remediate and less difficult to remediate groups had comparable levels of reading skill at the beginning of first grade. In contrast, the gain scores used as measures of response to first grade intervention accounted for significant variance on all of the reading outcome measures administered at the end of second and third grades (Vellutino et al., 2006). Accordingly, the authors posit that IQ and baseline achievement scores do not predict response to intervention, but more direct measures of response to intervention can strongly predict the intervention's effects (Vellutino et al., 2008). These findings are further supported by studies that examined characteristics of children who responded and did not respond to intervention.

In general, phonological awareness, rapid naming, verbal working memory, and oral language/vocabulary are believed to be the cognitive characteristics that can best differentiate responders from non-responders (Berninger et al., 2002; Case, Speece, & Molloy, 2003; Denton et al., 2013; Fletcher et al., 2011; Frijters et al., 2011; Vellutino et al., 2006). According to Stuebing et al. (2015), phonological awareness and rapid naming account for about 20-25% of the variance in reading outcomes after intervention, whereas IQ only accounts for 7% of the variance in response to reading intervention. Still, not to discount the importance of IQ, Fuchs and Young (2006) found that IQ explained 15% of the variance in reading comprehension when the reading intervention was comprehensive and 12% when the intervention was training in phonological awareness only. Thus, some researchers argue that IQ plays an important role in children's response to reading intervention (Fuchs & Young, 2006).

Nonetheless, Al Otaiba and Fuchs (2002) conducted a meta-analysis that identified children as responders or non-responders and found that a key characteristic of non-responding children was their phonological awareness. In fact, 70% of the reviewed studies found that phonological awareness was a crucial correlate of treatment unresponsiveness (Al Otaiba & Fuchs, 2002). They also found that phonological memory, rapid naming, intelligence, attention/behavior, and orthographic processing played a role in children's responsiveness to intervention. Similarly, Nelson, Benner, and Gonzalez (2003) conducted a meta-analysis and found that rapid naming, problem behavior, phonological awareness, alphabetic principle, memory, and IQ were the variables that influenced treatment response. In addition, Vaughn, Linan-Thompson, and Hickman (2003) propose that oral reading fluency can also distinguish between responders and non-responders. In their study of second-grade children at risk for RD, Vaughn et al. (2003) found that children who responded to intervention after 30 weeks and those who never responded to the 30-week intervention did not differ on oral reading fluency measures at pre-test, yet at post-test, responders read significantly faster than non-responders. Hence, there appears to be reading correlates (e.g., phonological awareness, rapid naming) that distinguish those children with reading difficulties who respond to intervention and those who do not.

Still, there is research that demonstrates that even those with the most severe reading difficulties who are considered non-responders can in fact improve their reading skills; however, whether they catch up to same-age peers is debatable (e.g., Denton et al., 2006; Lovett et al., 1994; Torgesen, 2001, 2002; Vaughn et al., 2012). For example, Torgesen et al. (2001) found that 8- to 10-year-old poor readers who made nearly no progress in reading over a 16-month period in their special education class made significant improvements in decoding and comprehension skills, after they were given two hours of daily one-to-one tutoring for eight weeks. In fact, 40% of the children were no longer in need of special education services within one year following the intervention (Torgesen et al., 2001). Still, although the students' reading accuracy and comprehension were in the average range after the follow-up period, their oral reading fluency rate remained severely impaired (Torgesen et al., 2001). Accordingly, though these difficult-to-remediate children made progress, they did not improve in all areas and catch up to their peers. Moreover, in a study of first- to third-grade children with persistent reading deficits, Denton et al. (2006) found that students' decoding, fluency, and comprehension significantly improved after intensive intervention; however, many students still remained below average. Similarly, for older children in middle school, intensive reading intervention may significantly improve their standardized scores, but they may continue to lack grade-level proficiency (Vaughn et al., 2012). Overall, much of this work on intervention studies with non-responders exemplifies that although some make progress after receiving additional intervention, many continue to struggle with reading, performing below what is expected for their age. Subsequently, in order to understand the overall reaching achievement of children with RD and with LRA, longitudinal research must be examined.

Longitudinal Trends in Reading Achievement

When thinking about the longitudinal trends in reading achievement, a concept called a Matthew effect is important to consider. A Matthew effect in education refers to the idea that the rich get richer and the poor get poorer (Stanovich, 1986, 1988; Walberg & Tsai, 1983). In other words, those who have the greatest academic skills at the start will get stronger throughout schooling, while those with the poorest academic skills will get weaker throughout schooling. First, the Matthew effect assumes that skill differences between poor and good readers remain stable over time (Morgan, Farkas, & Hibel, 2008b). Second, based on a Matthew effect, rates of growth should be slower for those with initially poor achievement and more rapid for those with higher achievement at the outset, which leads to an increasing achievement gap over time.

To date, there is inconsistent evidence as to whether a Matthew effect exists for reading (Protopapas, Parrila, & Simos, 2016). While some research suggests that the gap between poor and good readers widens over time (Bast & Reistma, 1998; Juel, 1988; McKinney, 1984;

McNamara, Scissons, & Gutknecth, 2011), other research suggests that this gap does not widen over time (Aarnoutse & van Leeuwe, 2000; Aunola, Leskinen, Onatsu-Arvilommi, & Nurmi, 2002; Catts et al., 2003; McCoach et al., 2006; Parrila et al., 2005; Phillips et al., 2002; Scarborough & Parker, 2003; Shaywitz et al., 1995). In regard to the research that suggests that the gap between high achievers and low achievers does not widen over time, Shaywitz et al. (1995) explains that those who are initially poor readers have a rate of reading development that is faster than high achievers; however, their reading skills never reach the level of their high-achieving counterparts. Similarly, Scarborough and Parker (2003) found evidence to suggest that the lower the initial reading skills of children, the faster their growth rate will be.

Other research also supports this idea that the gap between the achievement of children with RD versus those with typical reading remains stable over time or even narrows slightly time (Catts et al., 2003; Parrila et al., 2005; Pfost, Hattie, Dorfler, & Artelt, 2014; Phillips et al., 2002; Shaywitz et al., 1995). The narrowing of this gap over time has been termed the compensatory trajectory of reading development (Leppänen, Niemi, Aunola, & Nurmi, 2004; Parrila et al., 2005). According to this view, children with initially lower reading skills may catch up to their peers, whereas those with initially higher reading skills may display slower growth (Aarnoutse, van Leeuwe, Voeten, & Oud, 2001; Aunola et al., 2002; Leppänen et al., 2004; Parrila et al., 2005; Phillips et al., 2002). Some researchers propose that the period of development that is studied influences whether a compensatory trajectory or a Matthew effect is found. For example, Bast and Reistma (1998), McKinney (1984), and McNamara et al. (2011), all of whom found evidence for a Matthew effect, only studied reading development between kindergarten and third grade. Yet, when other researchers examined growth over a longer time period, for example from first to sixth grade, there was no evidence of a Matthew effect (e.g., Aarnouste & van Leeuwe, 2000; Phillips et al., 2002). Similarly, the way in which reading achievement is defined in research (e.g., individual reading-related processes versus broad reading achievement) may also contribute to whether a Matthew effect or a compensatory trajectory is found. Hence, based on this research, it is plausible that although children with RD likely never catch up to their non-RD peers in terms of achievement, they exhibit a faster rate of progress, which prevents them from falling even further behind (Huang, Moon, & Boren, 2014; Scarborough & Parker, 2003).

In contrast to this research that has found faster rates of improvement for children with RD, there is other research to suggest that the reading skills of children with RD improve at the same rate or even more slowly than their peers with typical reading skills. First, in their work to examine the effectiveness of special education, Morgan et al. (2010) found that when matched on characteristics related to the likelihood one will receive special education services (e.g., SES, gender, early reading skills), children receiving special education made statistically similar gains in reading achievement between third and fifth grade compared to those who did not receive special education services. Additionally, children in special education scored significantly lower than those not receiving services in fifth grade on a measure of reading achievement (Morgan et al., 2010). Interestingly, when covariates were not accounted for, those receiving special education services made significantly greater gains in reading scores between third and fifth grade compared to those who were not receiving services (Morgan et al., 2010). Therefore, in this study, results indicated that when matched to similar peers, children with disabilities did not grow at faster rates than their non-disabled peers and had lower reading achievement outcomes.

Although Morgan et al. (2010) used a large sample (n = 363 in special education; n = 5,995 not in special education), there are two relevant limitations to their work. First, researchers did not parse apart different disability categories, and the results for children with RD may differ, for example, from children with Autism Spectrum Disorder (ASD). As demonstrated by the research on the compensatory trajectory of reading development, it is possible that children with RD narrow the achievement gap (Leppänen et al., 2004; Parrila et al., 2005); however, given that this study grouped all disabilities together, one cannot determine if such a finding holds true. Second, this study only utilized two time points (i.e., third and fifth grade) to examine growth trajectories over time. Including more time points would allow for a better determination of the effectiveness of special education.

Despite the limitations of Morgan et al.'s (2010) study, there is further evidence to suggest that children with RD progress at slower rates in terms of reading development than children without RD. For example, in a study that used hierarchical linear modeling (HLM) to model reading and math growth trajectories of children with learning disabilities (LD), Morgan et al. (2011) found that compared to children without LD, children with LD had the lowest scores at each time point as well as the lowest scores as the end of fifth grade. The LD group also displayed a slower rate of growth in reading achievement between the spring of kindergarten and fifth grade (Morgan et al., 2011). In addition, by the end of fifth grade, the gap between the LD group and group of children without disabilities widened; however, there was not statistically significant evidence of a Matthew effect in reading (Morgan et al., 2011).

It is important to note that Morgan et al. (2011) utilized a small sample size of children with learning disabilities (n = 20), and researchers did not separate the LD group into the different types of LD (e.g., reading, writing, math). Thus, results of this study may not generalize specifically to children with RD. Additionally, this study only considered the growth trajectories of children with diagnosed RD, or those who were receiving special education services (Morgan et al., 2011). As a result, children with LRA who were not identified as RD were included in the control group of children without Individualized Education Programs (IEPs). Furthermore, given the time points that they used, the students' reading scores at kindergarten entry as well as differences in outcomes over a longer period of time (i.e., past fifth grade) were not examined (Morgan et al., 2011). Also of importance, the LD group was identified based on whether they had an IEP in the spring of kindergarten (Morgan et al., 2011); however, identification of learning disabilities in very early grades is often inaccurate (Rathvon, 2004; Scarborough, 1998).

Other research that examined special education outcomes in high-poverty schools implementing the Reading First initiative supports this finding of slower growth. For example, Sanford, Park, and Baker (2011) found that students in special education had slower growth rates on measures of oral reading fluency and reading proficiency. In addition, Katz et al. (2008) found that students with a SLD or a speech language impairment (SLI) had slower growth on measures of oral reading fluency, listening comprehension, and word analysis compared to children without disabilities from the beginning of second grade to the end of third grade. However, relative to their scores at the beginning of each year, the group with disabilities made significant progress in oral reading fluency each year. Still, it is important to note that both of these studies did not specifically examine children with RD; rather, those with RD were grouped with students with other types of disabilities.

In sum, the above described longitudinal research studies on children with RD indicate conflicting results. While there is some evidence to suggest that children with RD have faster reading growth rates than children without RD (Catts et al., 2003; Parrila et al., 2005; Phillips et al., 2002; Shaywitz et al., 1995), there is other evidence that suggests children with RD progress either at similar rates or more slowly than good readers (Katz et al., 2008; Morgan et al., 2010; Morgan et al., 2011; Sanford et al., 2008). Regardless of whether children have faster or slower growth rates, most of the evidence substantiates the idea that children with RD never catch up to their same-age peers in terms of their reading skills and frequently exhibit lower reading scores than their peers over time. Nonetheless, the work related to the reading growth of children with RD over time utilized small samples, did not examine extended periods of time (i.e., maximum of kindergarten through fifth grade), and did not always look at the trajectories specific to students with RD. Notably, none of these studies differentiated children with RD from children with LRA in their samples, which is necessary to do in order to understand whether there are differences between the two group's reading achievement over time.

Longitudinal Trends in Reading Achievement Between RD and LRA Groups

To start, it is important to point out that research that compares the reading achievement between children with RD and children with LRA is scarce. In fact, only three studies were found that claim to examine growth trajectories over time. Unfortunately, one study conducted by Jorm et al. (1986) assessed the reading and cognitive ability of 14 children with RD, 25 with LRA, and 414 normal readers between kindergarten and second grade, but the authors did not conduct any sort of analysis to allow for the comparison of trends in reading development over time.

Nonetheless, Francis et al. (1996) used individual growth curve analysis to examine the developmental lag hypothesis of RD between children with discrepancy-based RD and children with non-discrepant LRA. The developmental lag hypothesis suggests that children who are poor readers will eventually catch up to their peers who are good readers, such that there would be no differences in the end points of their reading and cognitive development (Francis et al., 1996). Francis et al. (1996) describe how previous research suggests that lag models may best explain the reading development of low achievers, if they have a better prognosis of catching up to their non-impaired peers, whereas a deficit model best explains the reading development of children with RD, it they do not catch up to their non-impaired peers. Further, if the developmental trajectories were found to be different between the two groups, then this would provide evidence that the two group represent different types of RD (Francis et al., 1996).

To conduct the study, Francis et al. (1996) longitudinally examined the reading performance (i.e., decoding, pseudoword decoding, comprehension) of children with RD (n = 29), children with LRA (n = 34), and children with average reading (n = 301) once per year between first and ninth grade. First, they hypothesized that for both groups, growth would be nonlinear, showing an initial pattern of rapid acquisition followed by slowing in the rate of change (Francis et al., 1996). After controlling for gender, race, and socioeconomic status (SES), results indicated that the RD and LRA groups were not significantly different from each other, in terms of starting points, rates of change, and rates of acceleration (Francis et al., 1996). These results are consistent with a deficit model (Francis et al., 1996). In other words, both the RD and LRA groups never caught up to their same-age peers with normal reading ability.

When children in the RD group who were not low achieving (i.e., performing above 25th percentile) were removed from the group, the RD group was found to have a significantly lower mean plateau than the LRA group (Francis et al., 1996). This means that while the two groups plateaued, or stopped progressing in terms of reading achievement at the same age (i.e., around 15 years), the RD group had a lower level of achievement at this plateau. This data further supports Shaywitz’s (1992a) findings that reading ability exists on a continuum, with those with RD representing the lower tail. Furthermore, the results suggest that children with RD and LRA progress at relatively similar rates and do not catch up to their peers.

Although Francis et al. (1996) utilized a strong longitudinal model, the sample size was relatively small, affecting both generalizability and power to detect statistically significant effects. Time was centered at eight years of age, so whether the two groups differed at kindergarten entry is unknown. Additionally, because the authors defined the groups using their own criteria, they did not specify which students received special education services or intervention in school. Thus, one cannot hypothesize about the influence of special education on the students' reading development over time.

The final longitudinal study that examined differences between children with RD and with LRA was conducted by O'Malley, Francis, Foorman, Fletcher, and Swank (2002). From kindergarten through second grade, researchers compared 29 children with LRA, five with RD based on discrepancy criteria, 20 who met both LRA and RD criteria, and 325 normal readers across a number of reading-related skills (O'Malley et al., 2002). The authors used HLM to estimate individual growth parameters and examine growth in letter-sound knowledge, phonemic awareness, rapid naming of objects, rapid naming of letters, visual-motor integration, perceptual discrimination, spelling, and word reading (O'Malley et al., 2002). The results indicated that the two groups did not differ in terms of intercept, slope, and acceleration on five of the eight reading-related skills (i.e., phonemic awareness, rapid naming of objects, perceptual discrimination, spelling, and word reading; O'Malley et al., 2002). That is, the two groups did not differ in skill-level at the beginning of first grade, had similar growth rates, and had similar acceleration of skill development (O'Malley et al., 2002).

For the other three reading-related skills, the two groups significantly differed on only one parameter for each of the skills. For letter-sound knowledge, the RD group demonstrated faster growth than the LRA group; however, this difference in the slope parameter was not statistically significant (O'Malley et al., 2002). For rapid naming of letters, the RD group exhibited significantly less deceleration over time, and for visual-motor integration, the RD group had a higher mean-level of performance at the beginning of first grade (O'Malley et al., 2002). Overall, the results of this study suggest that there are minimal differences in the early reading-related skills of children with RD and with LRA between kindergarten and second grade.

Nonetheless, this study was limited by a number of factors. First, the sample size was small, so power to detect significant results may be inadequate. Second, this study did not look at an overall reading achievement composite; rather, only the reading-related skills were examined between kindergarten and second grade. Third, though researchers collected data at the kindergarten time point, they used first grade as the intercept. Thus, one cannot conclude that the two groups of children had similar reading-related skills at the start of kindergarten. Next, this study did not extend past second grade, so how similar or different the two groups looked in later grades is unknown. Lastly, because the authors determined group categorization based on their own criteria, whether participants received special education services was not discussed.

Overall, the two studies conducted by Francis et al. (1996) and O'Malley et al. (2002) provide evidence that children with RD look more similar than different compared to children with LRA. However, there are a number of limitations to each of these studies and, thus, could be strengthened by future research. Perhaps, one of the biggest improvements that could be made to this research is to include information on special education eligibility. By only providing information about how the RD and LRA groups were defined for the research study (e.g., RD = achievement score was 1.5 standard errors below score predicted by IQ; LRA = achievement score below the 25th percentile) and not how they were classified in school, conclusions about the effectiveness of the services they received in school could not be made. In addition, larger, nationally representative samples are needed in order to generalize the results of the research. Finally, using data over a longer period of time (e.g., kindergarten through eighth grade) would provide more information about whether the reading growth trajectories of students with RD and students with LRA differ over time.

In many of the studies described above, researchers recognized the importance of considering factors aside from reading scores that affect reading development over time; thus, these additional factors were often used as control variables. For instance, Francis et al. (1996) controlled for gender, race, and SES. Morgan et al. (2011) controlled for age, SES, race, school type, and kindergarten retention. Unfortunately, from the O'Malley et al. (2003) study, it is unclear if they controlled for any additional variables. Other studies discussed above used covariates such as race, sex, age, grade, language status, SES, reading level, learning-related behaviors, preschool experiences, birth weight, number of siblings, etc. (Katz et al., 2008; Morgan et al., 2010; Sanford et al., 2008). In all, while there is some variability in which variables the researchers chose to control for, most researchers included age, gender, race, and SES at minimum. These important covariates are discussed further below.

Factors that Influence Reading Achievement

Accordingly, the current study recognizes the importance of not only within-child factors but also environmental factors that influence children's interactions and subsequent reading development. The main environmental variable that are considered in this study is SES at the kindergarten assessment. There are also three within-child variables in addition to reading achievement that are considered. Such variables are age at the fall kindergarten assessment, gender, and race.

However, given the purpose and methodology of this study, there are some environmental and within-child variables that are not examined, but that it is not say that they do not play a role in children's reading development. These variables are described below, followed by a discussion of the additional variables that are considered in the current study.

Variables Not Controlled in the Current Study

Possible covariates that are not included in the study are IQ, language, motivation, and early reading experiences (e.g., preschool; home-literacy environment).

IQ. Although there is much research to suggest that the IQ-achievement discrepancy is an invalid way to identify students with RD, some researchers still argue that IQ plays an important role in predicting reading ability (Fuchs & Young, 2006; Tiu, Lee, Thompson, & Lewis, 2003). The current study, however, does not control for IQ, because the extant dataset does not have a measure of IQ. Nonetheless, since the data for this study were collected before and in the early years of the RTI movement, many of the students may have been found eligible or ineligible for special education on the basis of the IQ-achievement discrepancy. Still, to ensure that the students with significantly lower IQs are not included in the study, children with neurological conditions and intellectual disability are excluded.

Language. There are two aspects of language that are important to consider. One is language ability and the other is the language the child speaks. Regarding language ability, RD is considered a language-based disorder that affects not only a child's ability to read but also his/her ability to speak and spell (Shaywitz, 2003). Children with RD have also been found to have impairments in semantics and syntax (Gallagher, Frith, & Snowling, 2000; Snowling, Gallagher, & Frith, 2003). Given that language and early reading skills are intertwined and that there are a number of factors that influence a child's language development (Holm, Crosbie, &

Dodd, 2007; Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010), controlling for early language ability would remove differences in reading achievement scores that may be due to language development. Nevertheless, language ability is not controlled in the current study, as the extant dataset has no measure of language ability.

Second, related to the relationship between the child's English language status and reading, there is research to suggest that children who are English Language Learners (ELLs) have lower reading achievement compared to native English speakers (see August & Shanahan, 2006 for review). However, there is much research to suggest that the lower reading achievement of ELLs is more complicated than just differences in their spoken language; rather, much of the difference between the reading achievement of ELLs and native-English speakers can be captured by race/ethnicity and/or SES. Approximately 72% of ELL students are Hispanic (Planty et al., 2008), and at least 75% of the ELL students across the United States are eligible for a free or reduced-price school lunch program (Zehler et al., 2003). Even further, Kieffer (2008) examined the reading growth trajectories of ELLs compared to native English speakers from kindergarten through fifth grade and found that once the child's ethnicity and SES were taken into account, the growth trajectory was not affected by whether the child spoke a language other than English in the home. Thus, in order to maintain parsimony in the current model, the child's English language status is not controlled, but other related variables such as SES and race are covariates in this study.

Motivation. Children with reading difficulties are often unmotivated to read (Morgan, Fuchs, Compton, Cordray, & Fuchs, 2008c) and are less likely to practice reading (Guthrie, Wigfield, Metsala, & Cox 1999). Therefore, low motivation is believed to be a key factor underlying students' long-term reading difficulties (Lepola, Poskiparta, Laakkonen, & Niemi, 2005; Sideridis, Morgan, Bostas, Padeliadu, & Fuchs, 2006). Morgan and Fuchs (2007) found that there is a reciprocal relationship between children's reading skills and motivation to read, suggesting that both reading and motivation need to be the target of early intervention. In the current study, motivation is not used a covariate in order to retain the variance that motivation contributes to reading achievement scores.

Early reading experiences. A child's early reading experiences including whether they attend a high-quality preschool as well as their home-literacy environment can influence reading development. According to Snow et al. (1998), a preschooler who has fewer opportunities for acquiring knowledge pertaining to books and reading in the home is at a higher risk for developing reading difficulties than a child whose home allows for a richer literacy environment. Similarly, children with more preschool experience have been found to have higher achievement scores in second grade (Pianta & McCoy, 1997). Not surprisingly, both home-literacy environment and early childcare experiences are related to SES, as children from economically disadvantaged homes are less likely to have enriching home-literacy environments and attend preschool (Rathbun & Zhang, 2016; van Steensel, 2006). Therefore, in the present study, early reading experiences are not controlled, because SES is a covariate.

Variables Controlled in the Current Study

Covariates included in the study were gender, SES, race, and age.

Gender. Research suggests that there are gender differences in reading development. Girls have been found to develop reading skills more quickly than boys (Scheiber, Reynolds, Hajovsky, & Kaufman, 2015; Wolf & Gow, 1986). Girls have also been found to score higher than boys in reading, and a higher percentage of girls reach reading proficiency (NCES, 2002). Thus, it is possible that differences in reading achievement can, in part, be attributed to gender.

In addition, reading disabilities have been shown to be more prominent in boys than in girls (Rutter et al., 2004). Even when taking severity into account, boys tend to show more reading difficulties than girls (Olson, 2002). However, research in support of these findings is typically based on school- or clinic-referred samples. In schools, greater RD diagnoses may result from a greater number of boys being referred for special education services (Anderson, 1997). Furthermore, results from epidemiological studies suggest that there are small to no significant gender differences in RD. For example, an epidemiological study of RD in Connecticut found a male-to-female ratio of 1.2:1, which further demonstrates the equal prevalence of RD across boys and girls (Shaywitz, Shaywitz, Fletcher, and Escobar, 1990). Accordingly, depending on the type of research and the way in which samples are accrued, the ratio of boys to girls with reading difficulties may vary, and as a result of this variation, it is important to control for the effects of gender on reading achievement.

Socioeconomic status (SES). SES is an indicator of a child's social status, and previous research has identified a relationship between SES and reading achievement (Denton & West, 2002; Dickinson & McCabe, 2001; Molfese, Modglin, & Molfese, 2003). Generally, poverty has consistently been found to have negative effects on reading achievement (Denton & West, 2002; Lee, Grigg, & Donahue, 2007; Morgan et al., 2008b). Preschool measures of SES and home environment predict reading scores between 8 to 10 years of age (Molfese et al., 2003). More specifically, children from economically disadvantaged homes acquire oral language skills more slowly, exhibit delayed letter recognition and phonological sensitivity, and are at risk for reading difficulties (Hoff, 2013; Whitehurst & Lonigan, 1998). Additionally, reading outcomes have been found to be significantly related to certain factors of family context, such as home-literacy environment, number of books owned, parental distress, and receipt of center-based cared

(Aikens & Barbarin, 2008). Thus, SES is positively correlated with children's reading achievement, as those parents with higher SES tend to be able to provide their children with more resources, academic opportunities, and verbally rich environments than those parents with lower SES (Davis-Kean, 2005).

Furthermore, children with higher SES are more likely to be diagnosed with RD, whereas those with lower SES are more likely to be considered poor readers (Siegel & Himel, 1998). There is also research to suggest that the rate of special education placement for RD is significantly greater for children from low SES backgrounds than those not experiencing low SES (Blair & Scott, 2002). Consequently, in the schools, SES likely affects whether a child will receive special education services. Based on these findings, it is necessary to control for the differential effects of SES on reading achievement.

More specifically, it is important to control for the effects of early SES, specifically at the kindergarten time point. There is now a copious amount of research to suggest that early economic disadvantage and adversity can have a significant effect on neurological development and learning (Blair & Raver, 2012; Brooks-Gunn & Duncan, 1997; Hair, Hanson, Wolfe, & Pollak, 2015; Johnson, Riis, & Noble, 2016; Yoshikawa, Aber, & Beardslee, 2012). Further, children living in poverty tend to have a physiological response to stress, indicated by increased levels of stress hormones, such as cortisol. Accordingly, researchers have found that the executive functions (e.g., working memory, inhibitory control, mental flexibility) and IQs of three-year-old economically disadvantaged children are negatively correlated with cortisol levels at 7-, 15-, and 24-months of age (Blair et al., 2011). This suggests that the relationship between poverty and the physiological response to stress (i.e., increased cortisol levels) are associated with intelligence and executive functioning, even before a child enters formal schooling.

Researchers have also found that children from lower-SES families perform worse on executive function measures at 54 months of age, and although low-SES and high-SES groups develop executive functions at the same rate, the gap in executive function performance between the two groups does not narrow over time (Hackman, Gallop, Evans, & Farah, 2015). fMRI studies have also provided evidence of neural differences in brain regions associated with language and executive functions between infants from high- and low-SES backgrounds (Tomalski et al., 2013). These findings further point to the importance of the effects of early economic disadvantage on later cognitive development. In addition to executive functions and IQ, SES and poverty have a strong influence on children's language skills, especially during early childhood (Demir & Küntay, 2014; Noble, Norman, & Farah, 2005). Therefore, it is important to control for SES at the earliest possible time point in order to (1) mitigate differential effects of SES on reading achievement, and (2) examine how early SES plays a role in later reading achievement in students with RD and LRA.

Race/ethnicity. Children from racially/ethnically diverse groups, such as African American and Hispanic children, are disproportionately over-represented in special education (Coutinho & Oswald, 2000; Hosp & Reschly, 2003; Oswald, Coutinho, Best, & Singh, 1999), and federal and state legislation have not been effective in remediating this disproportionality (Albrecht, Skiba, Losen, Chung, & Middleberg, 2012). Over-representation in special education of racial minorities has also been documented specifically for students with learning disabilities (Coutinho, Oswald, & Best, 2003).

Even though children of diverse racial groups are over-represented in special education, they have been found to have lower levels of achievement. For example, in a longitudinal study of reading achievement, Morgan et al. (2008b) found that African American, Hispanic, and

Asian children had slower rates of reading growth than White children, and African American children experienced the greatest increase in the gap relative to White children. Furthermore, Hispanic children entered school with significantly fewer reading skills than White children (Morgan et al., 2008b). As a result, it is important to control for the effects of race on reading achievement.

Age. Over recent years, the kindergarten curriculum had become increasingly focused on academics (Cosden, Zimmer, & Tuss, 1993); however, many parents are delaying kindergarten entry. Nonetheless, research on the effects of delaying kindergarten on achievement has been inconclusive. Some studies have found higher academic achievement for older kindergarten students (Stipek & Byler, 2001), while other research suggests that kindergarten age is not a good predictor of later achievement (Morrison, Griffith, & Alberts, 1997).

Still, research indicates that reading is a developmental process that progresses with age, although there is variation among individual children (Snow et al., 1998). Given that children may be of different ages at kindergarten entry, older children may have more well-developed reading skills compared to younger children as result of their advanced age. Lastly, one study found that younger children were more likely to be classified as having RD while older children were more likely to be classified as low achieving (Siegel & Himmel, 1998). Accordingly, it is important to control for age in order to remove any variance in the reading achievement scores related to developmental maturity.

Purpose of the Present Study

In recent years, reading development and reading instruction have been at the forefront of federal policies and research. The ultimate goal of these initiatives is to prevent reading failure and help all children to acquire strong reading skills in early elementary school via high-quality, evidence-based instruction. Despite these efforts, there are a number of children who struggle with learning how to read. Some of these struggling readers will be diagnosed with RD and receive special education services, while others will continue to struggle with reading but will not be identified to receive such services.

Historically, there has been great debate about whether children with RD are distinguishable from children with LRA. Most research in this area suggests that children with RD and LRA are not distinguishable across the majority of neurocognitive measures (Shaywitz et al., 1992b; Ysseldyke et al., 1982), although there is some research that has found differences (Fuchs et al., 2002). Nonetheless, there is substantial research to suggest that deficits in phonological awareness are at the core of reading problems for both children with RD and those with LRA (Fletcher et al., 1994; Stanovich & Siegel, 1994; Stuebing et al., 2002). Even though differences between the two groups across most standardized measures are not usually found, researchers have identified groups of children who respond differently to intervention, such that some children never catch up to their same-age peers (Denton et al., 2006; Lovett et al., 1994; Torgesen et al., 2002; Vaughn et al., 2012). It is these children, who are least likely to respond to intervention, that are arguably those with the most severe reading difficulties, requiring the most substantial and intensive intervention.

Although it may be possible to distinguish children with RD from those with LRA based on their response to intervention, limited research exists regarding the differences in growth trajectories between the two groups. Research that has examined reading achievement longitudinally typically compares either children with RD to children without RD or children with many types of disabilities to children without disabilities. In general, this body of work is limited by small sample sizes, narrow time points, and groups where RD and LRA students cannot be distinguished.

More specifically related to the differences in the growth trajectories between children with RD and children with LRA, only two research studies have addressed this question. Both studies indicated that there were no differences in the starting points, rates of growth, and outcomes between the two groups (Francis et al., 1996; O'Malley et al., 2002). Unfortunately, these studies only identified students for each group using research-based criteria, so their school-based eligibility is unknown. Examining the growth trajectories between students with RD and students with LRA using a larger sample, over a longer time period, and based on school diagnostic eligibility will allow for a better understanding of differences in the reading achievement between these two groups. Additionally, the use of more advanced statistical methods that can account for the relationship between other factors that may influence reading achievement as well as examine trends over time will add to the literature in this area.

The current study will test a longitudinal model examining the differences in reading growth trajectories from kindergarten through eighth grade between students who received special education services between first and fifth grade for RD and non-identified children with LRA between first and fifth grade. Latent growth modeling (LGM) will be used to empirically test the proposed model and determine if there are differences in reading achievement start points, patterns of change, and outcomes, while controlling for important demographic factors (i.e., age, gender, SES, race/ethnicity; see Figure 3). Primary analysis will use multiple group LGM to simultaneously fit the model to the populations of interests (i.e., RD, LRA) and determine whether the growth trajectories differ between the two groups.

Utilizing a large, representative sample of children who entered kindergarten in the fall of 1998, this study seeks to better understand whether children with RD and children with LRA differ in their reading achievement over time. Examination of these groups over an extended period of time using school-based classification addresses previous methodological flaws. Results of this study will contribute to an improved understanding of whether children with RD who receive special education are different in terms of reading development from students with similar LRA who do not receive such services. Additionally, the results may further help researchers to conceptualize diagnostic criteria for RD in schools, clinical settings, and research. Lastly, the results are intended to inform researchers and policymakers as they seek to develop the most effective, evidence-based instruction and intervention to support the learning of both students who receive special education services and students who do not receive such services but also struggle in reading.

Research Questions

Question 1

To what extent do SES, gender, race, and age predict whether a student will be identified as having either RD or LRA between first and fifth grade?

Question 2

How does initial reading achievement in the fall of kindergarten compare between children who received special education services and non-identified children with LRA with SES, gender, race, and age controlled?

Question 3

How does the pattern, or shape, of reading growth from kindergarten through eighth grade compare between children who received special education services and non-identified children with LRA with SES, gender, race, and age controlled?

Question 4

How do reading achievement outcomes in the spring of eighth grade compare between children who received special education services and non-identified children with LRA with SES, gender, race, and age controlled?

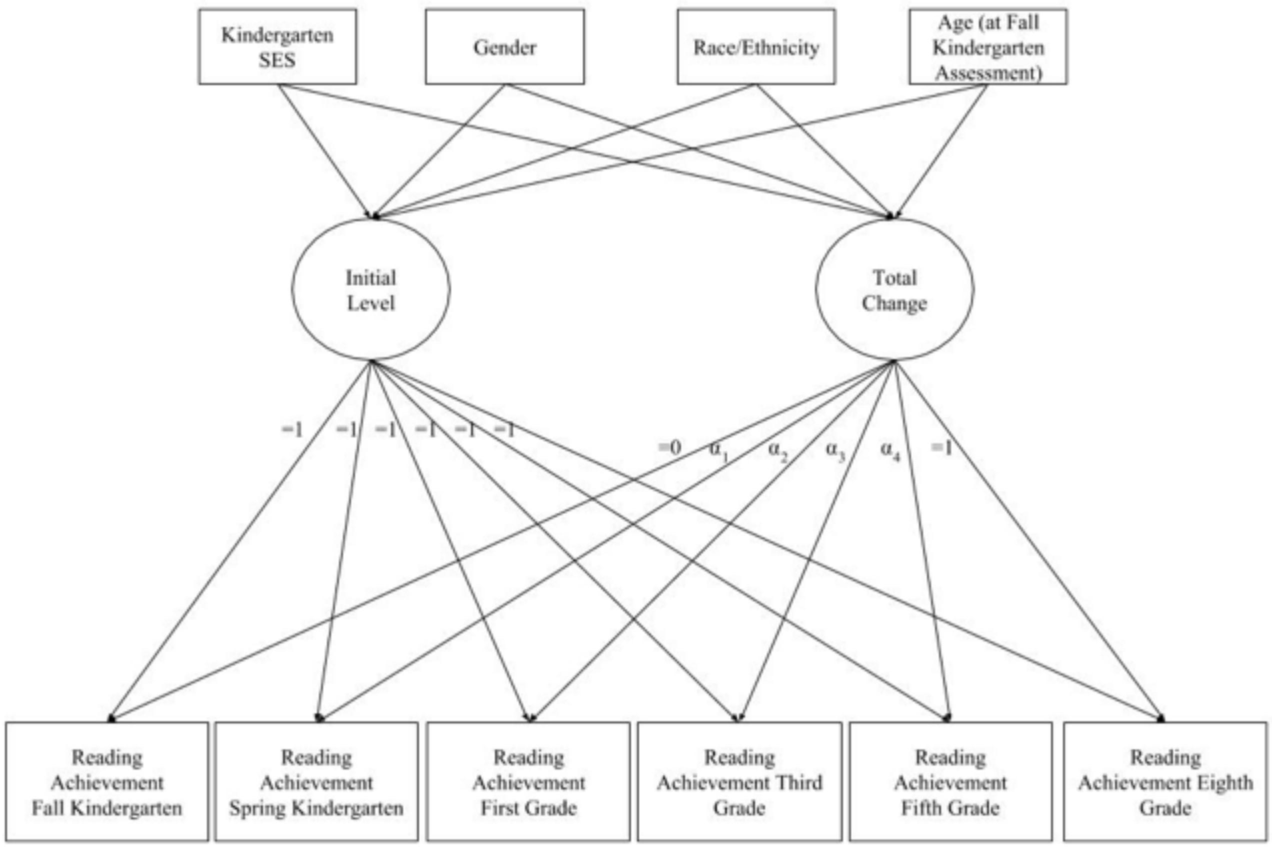

Figure 3. Simplified Latent Basis Growth Model for the Current Study. $\alpha_{1\text{-}4}$ = freely estimated basis coefficients.

CHAPTER 3
METHOD

Study Design

For this study, data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999 (ECLS-K) was be used to examine trends in reading achievement from kindergarten through eighth grade between children with an identified RD and non-identified children with LRA. These data were collected by the U.S. Department of Education's National Center for Education Statistics (NCES) within the Institute of Education Science (IES), in which a nationally representative sample of children were followed from the fall of kindergarten through the spring of eighth grade. This current study took advantage of the longitudinal data to understand differences in the trajectories of reading achievement of children receiving special education services for RD versus non-identified students with LRA. The ECLS-K user manuals (Tourangeau et al., 2004; Tourangeau et al., 2002; Tourangeau et al., 2009; Tourangeau et al., 2006) and psychometric reports (Najarian, Pollack, & Sorongon, 2009; Pollack, Atkins-Burnett, Najarian, & Rock, 2005a; Pollack, Atkins-Burnett, Rock, & Weiss, 2005b; Rock et al., 2002) contain information about data collection, samples, and variables of interest, all of which are described below.

Data Collection Waves

The ECLS-K utilized multiple methods and multiple sources to collect in-depth data beginning in kindergarten and ending in eighth grade on variables related to the children's development, school experiences, and family experiences. The longitudinal dataset includes a total of 21,410 students. Please note that all sample sizes are rounded to the nearest 10 for data security purposes. The samples at each wave of data collection are described below.

The first and second waves, or base year, of data collection occurred during the fall and spring of 1998-1999. The fall kindergarten sample included 21,390 children, and the spring kindergarten sample included 22,810. The increase in the number of children between the fall and spring kindergarten data collection waves resulted from the inclusion of an additional 1,430 children who were sampled from schools that originally refused participated but agreed by the spring kindergarten data collection wave.

The third and fourth waves of data collection occurred during the fall and spring of first grade in 1999-2000. The fall first-grade data collection did not include the full kindergarten sample; rather, it was limited to a 30% subsample of schools (i.e., 27% of the eligible base-year children) in order to better analyze the effects of the loss of learning over the summer. The fall first-grade sample included 6,510 children. The spring first-grade sample was then expanded to include all children who were eligible during the base year assessment (n = 21,360). During the spring of first grade, all children in a random 50% subsample of base-year schools were flagged to be followed if they transferred from their base-year school at any point in the future. Additionally, in the spring of first grade, a half-open interval sampling procedure was used in the same 30% subsample of schools that was used in the fall first-grade data collection to "freshen" the spring sample. In other words, children who were not in kindergarten in the United States during the 1998-1999 school year (e.g., immigrants, children who repeated first grade in 1999-2000) were added to the spring first-grade sample. A total of 165 students were added to the sample based on the freshening procedures. These children were not included in the current study's sample as a result of missing data in kindergarten.

The final three waves of data collection occurred in the spring of third, fifth, and eighth grade. The fifth wave of data collection occurred in the spring of 2002 when most children were in third grade (*n* = 21,360). The sixth wave of data collection occurred in the spring of 2004 when most children were in fifth grade (*n* = 16,140). In fifth grade, 5,210 children were excluded from data collection as a result of ineligibility in a previous wave, movers who were not subsampled to be followed, hard-to-field cases, hard refusals, and children with neither first nor third grade data. The seventh and final wave of data occurred in the spring of 2007 when most of the children in the sample were in eighth grade (*n* = 12,130). Because the sample was not freshened after the first-grade wave, data from the ECLS-K eighth-grade wave are not representative of all eighth-graders in 2007, but are representative of the population cohort. After comparing the weighted population of the ECLS-K eighth-grade sample to the weighted population of the 2006 Current Population Survey from the U.S. Department of Labor, results suggest that the ECLS-K data represent about 80% of all U.S. eighth-graders during the 2006-2007 school year. Examples of those not represented in the eighth-grade sample include children who started kindergarten before the fall of 1998 and were retained in a later grade, children who immigrated to the United States after first grade, and children who were home-schooled until after first grade. Please see Table 1 for an overview of the ECLS-K data collection sample.

Data Collection Methods

During all waves of data collection, interviews, questionnaires, and direct child assessments were used to gather information from parents, teachers, school administrators, and the children themselves. Computer-assisted interviewing (CAI) was used to gather information from parents predominantly over the phone, but a small percentage of interviews were conducted in person. Interviews were conducted primarily in English, but bilingual interviewers or interpreters were provided for parents whose primary language was not English. For non-English speaking parents, a hard-copy questionnaire was used to record interview answers, which were then input into the CAI. Average parent interviews across the waves of data collection ranged from approximately 44 to 65 minutes.

General education teachers, special education teachers, and school administrators completed self-administered questionnaires across all waves of data collection. General education teachers answered questions related to their classroom characteristics, classroom instruction and teaching methods, and specific questions about the sampled students in their classrooms. Special education teachers answered questions about their professional background and experience as well as the special educations services provided to the sampled students. School administrators answered questions about the school's student body, teachers, school policies, and administrator characteristics.

Students completed direct cognitive assessments using pencil and paper as well as CAI. In kindergarten and first grade, the direct cognitive assessment consisted of questions in three subject areas (i.e., language and literacy, mathematical thinking, and general knowledge), and it was conducted in a one-on-one testing setting. In third through eighth grade, the direct cognitive assessment consisted of questions in the same areas as earlier grades, but general knowledge was changed to science. In eighth grade, the direct cognitive assessment was administered in a group setting, as opposed to a one-on-one setting. In addition, starting in third grade and continuing through eighth grade, students completed a socio-emotional questionnaire.

Data Collection Procedures

Prior to the collection of data for each wave, all field staff, including field supervisors, assessors, and interviewers participated in extensive training. Prior to the fall kindergarten wave of data collection, field staff were selected based on their previous experience in working on national, large-scale studies (i.e., NAEP, TIMSS). The selected field staff completed eight hours of home study training on the direct child assessments, field procedures, and computer skills along with an additional eight hours of training in interviewing techniques. Subsequently, field supervisors participated in a two-day in-person training, and assessors attended a five-day in-person training. In between each wave of data collection, additional training was provided to field staff either via home study materials or in-person training. Additionally, prior to both the third-grade and the fifth-grade data collection, field supervisors and assessors completed certification exercises (i.e., written assessment, observation) to ensure that the direct child assessments would be administered in a standardized manner. For the third-grade data collection, 98.2% of trainees passed the certificate procedures, and 1.8% were required to complete both remedial training and an additional observation in the field before being approved. For the fifth-grade data collection, 99.1% of trainees passed the certificate procedures, and 0.9% were required to complete both remedial training and an additional observation in the field before being approved.

Table 1

Overview of Original ECLS-K Data Collection Sample

Data Collection Wave	Unweighted Sample Size[a]	Unweighted Number of Respondents[a]	Data Collection Period	Sample
Kindergarten F	21,390	19,680	Fall 1998	Full sample
Kindergarten S	22,810	20,580	Spring 1999	Full sample
1st Grade S	6,510	5,420	Fall 1999	30 percent subsample
1st Grade S	21,360	17,320	Spring 2000	Full sample plus freshening
3rd Grade S	21,360	15,310	Spring 2002	Full sample
5th Grade S	16,140	11,820	Spring 2004	Full sample
8th Grade S	12,130	9,730	Spring 2007	Full sample

Note. F = fall; S = spring. [a]Unweighted sample size and unweighted number of respondents represents unweighted sample rounded to nearest 10.

ECLS-K Original Sample

The ECLS-K used a stratified, clustered, multistage probability sampling design to select a nationally representative sample of children attending kindergarten in 1998-1999. The sampling procedures included three stages: primary sampling units (PSUs; i.e., geographic areas consisting of counties or groups of counties); second-stage units (i.e., schools within sampled PSUs); and third-stage units (i.e., students within schools).

PSUs were developed from 1990 county-level population data and 1994 population estimates of five-year-olds by race/ethnicity. If a PSU did not have at least 320 five-year-olds, then it was collapsed with an adjacent PSU. After collapsing, 1,335 PSUs remained in the final PSU frame. The selection of PSUs took into account the amount of oversampling of Asian and Pacific Islanders needed to meet the nationally representative sample size criteria. Asian and Pacific Islanders were the only subgroup that was oversampled. Subsequently, 100 PSUs were selected for the ECLS-K.

In second-stage sampling, public and private schools offering kindergarten programs were selected. After obtaining data to create the public and private school sampling frame, schools within the chosen PSUs were identified. The school frame included 18,911 public schools and 12,412 private schools. Public schools within the frame had at least 24 kindergarteners and private schools had at least 12 kindergarteners, and if a school had less than those amounts, the schools were clustered together. Schools were selected based on the probability proportional to the size of the PSU. In other words, the number of schools designated to each PSU was proportional to the weighted measure of the size of the PSU, and if a PSU was so small that a school may not have been allocated to it, at least one school was imposed on that PSU. In all, 1,280 schools were selected for the ECLS-K (i.e., 934 public schools and 346 private schools). Of these sample schools, 75% of public schools and 93% of private schools cooperated for purposes of the study.

In third-stage sampling, two independent sampling strata were created within each school. One stratum was for Asians and Pacific Islanders, and the other stratum was for all remaining students. Within each stratum, students were selected using equal probability systematic sampling. When a twin was sampled, both twins were included as a unit in the sample. At most schools, the target number of children to sample was 24, and in some schools, the goal of oversampling Asian and Pacific Islanders could not be met. Upon identifying sampled children, parent contact information was obtained from the schools, so parents/guardians could provide consent for their child’s participation and for the parent interview. Please see Table 2, Table 3, and Table 4 for a breakdown of the original ECLS-K sample at each wave of data collection by race/ethnicity, gender, and age, respectively. Over time, the race/ethnicity and gender characteristics of the sample remained roughly the same.

Table 2

Breakdown of Original ECLS-K Sample at Each Wave of Data Collection by Race/Ethnicity

Wave[a]	Race/Ethnicity							
	White/ Non-Hispanic	African American/ African-American	Hispanic, Race Specified	Hispanic, No Race Specified	Asian, Non-Hispanic	Native Hawaiian or Other Pacific Islander	Native American or Alaskan Native	More Than One Race
K1	55.2%	15.1%	8.3%	9.4%	6.4%	1.0%	1.8%	2.4%
1st	56.4%	14.3%	8.2%	9.1%	6.3%	1.1%	1.8%	2.4%
3rd	56.6%	13.0%	8.9%	9.2%	6.5%	1.2%	1.8%	2.6%
5th	57.0%	11.4%	9.3%	9.8%	6.9%	1.3%	1.9%	2.4%
8th	60.9%	10.3%	8.5%	9.0%	5.7%	1.1%	2.1%	2.3%

Note. K1 = fall of kindergarten. [a]Wave represents the data collected at designated grade.

Table 3

Breakdown of Original ECLS-K Sample at Each Wave of Data Collection by Gender

Wave[a]	Gender	
	Male	Female
K1	10,870 (51.1%)	10,380 (48.8%)
1st	8,950 (51.2%)	8,530 (48.8%)
3rd	7,810 (51.0%)	7,490 (49.0%)
5th	5,990 (50.7%)	5,830 (49.3%)
8th	4,930 (50.7%)	4,800 (49.3%)

Note. K1 = fall of kindergarten. All values of *n* rounded to nearest 10. [a]Wave represents the data collected at designated grade.

Table 4

Breakdown of Original ECLS-K Sample at Each Wave of Data Collection by Age

Wave[a]	Age of Study Child (Months)	
	Mean	SD
K1	65.79	14.80
K2	74.63	4.87
1st	86.84	4.84
3rd	111.06	4.78
5th	134.93	4.93
8th	171.25	6.69

Note. K1 = fall of kindergarten. [a]Wave represents the data collected at designated the grade.

Included and Excluded Samples

Given that the population of interest in the current study is children with reading difficulties, many participants in the ECLS-K dataset were excluded from the final sample. Ninety-five percent of the entire ECLS-K sample (n = 20,330) were excluded, producing a final sample consisting of 1,080 children. To better understand the differences between the included and excluded samples, comparison tests were conducted to determine if there were significant differences in both categorical variables and continuous variables of interest (see Table 5 and Table 6). Not surprisingly, when comparing the included sample to the entire ECLS-K excluded sample, there were significant differences. Specifically, differences in predictor variables (i.e., gender, race/ethnicity, SES, age at fall kindergarten assessment) were observed, because such variables are hypothesized to play a role in academic achievement and special education eligibility. Significant differences in the reading assessment IRT scale score outcome variable were also observed, because the included sample is defined by special education eligibility and poor reading achievement.

To further examine differences between the included and excluded samples, comparison tests were conducted for each group in the study. Table 7, Table 8, Table 9, and Table 10 compare the categorical and continuous variables of interest for the RD and LRA groups, respectively. Those in the excluded groups for this comparison met exclusionary criteria described below. For both the RD and LRA, there were no significant differences in gender or race/ethnicity between included and excluded samples. The included and excluded samples of the RD group significantly differed in SES, in which the included RD group had a lower SES. The included and excluded samples of the LRA group significantly differed in the age at the fall kindergarten assessment, in which the included sample was on average 1.4 months younger. Additionally, the included and excluded samples of both the RD and LRA groups had significantly different IRT scale scores on the reading assessment at certain waves of data collection. When significant, the included participants scored higher than the excluded participants. Although there were statistically significant differences between the included and excluded groups, not all of the differences are likely to be clinically significant. Those in the excluded group may not only have missing data but also may have met the exclusionary criteria of having a more severe special education eligibility (e.g., intellectual disability) and, therefore, scored lower on the direct reading assessment. It was originally intended that sample weights would be applied to all analyses to account for non-response biases and maintain generalizability of the final results to the population. In other words, the use of sample weights would have helped account for those with missing data due to non-response to yield representative estimates for the parameters of interest.

Table 5

Comparison of Categorical Variables for Included and Excluded Samples

Characteristics	Included Sample N (%)	Excluded Sample N (%)	Chi-Square (*df*)
Total	1,080 (5.0%)	20,330 (95.0%)	
Child Gender			8.96 (<*10*)**
Male	600 (55.6%)	10,340 (50.9%)	
Female	480 (44.4%)	9,960(49.1%)	
Child Race/Ethnicity			429.69 (*10*)***
White/Non-Hispanic	350 (32.8%)	11,450 (56.5%)	
African American/African-American	200 (18.7%)	3,040 (15.0%)	
Hispanic, Race Specified	150 (13.8%)	1,630 (8.1%)	
Hispanic, No Race Specified	230 (21.0%)	1,800 (8.9%)	
Asian, Non-Hispanic	50 (4.8%)	1,320 (6.5%)	
Native Hawaiian or Other Pacific Islander	20 (1.4%)	210 (1.0%)	
Native American or Alaskan Native	60 (5.9%)	320 (1.6%)	
More Than One Race	20 (1.6%)	500 (2.5%)	

Note. All values of *n* rounded to nearest 10. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 6

Comparison of Continuous Variables for Included and Excluded Samples

Variables	Included Mean (*SD*)	Excluded Mean (*SD*)	Independent Samples *t*-test (*df*)
Kindergarten SES	-0.48 (*0.69*)	-0.03 (*0.80*)	20.20 (*20140*)***
Age at Fall Kindergarten Assessment	67.78 (*4.70*)	68.50 (*4.48*)	5.14 (*19110*)***
Reading Assessment IRT Scale Score			
K1 Grade	28.45 (*4.11*)	36.89 (*10.24*)	23.25 (*8130*)***
K2 Grade	35.86 (*5.62*)	48.72 (*14.11*)	26.74 (*9040*)***
1st Grade	52.72 (*9.89*)	82.36 (*23.09*)	39.29 (*9210*)***
3rd Grade	90.07 (*15.52*)	134.02 (*25.09*)	55.25 (*9250*)***
5th Grade	116.78 (*17.94*)	156.33 (*23.57*)	52.79 (*9220*)***
8th Grade	138.18 (*25.86*)	175.26 (*24.84*)	45.26 (*9220*)***

Note. K1 = fall of kindergarten; K2 = spring of kindergarten. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 7

Comparison of Categorical Variables for Included and Excluded RD Samples

Characteristics	Included Sample N (%)	Excluded Sample N (%)	Chi-Square (*df*)
Total	300	430	
Child Gender			2.26 (*1*)
Male	190 (64.0%)	300 (69.3%)	
Female	110 (36.0%)	130 (30.7%)	
Child Race/Ethnicity			6.85 (*7*)
White/Non-Hispanic	180 (60.3%)	230 (52.3%)	
African American/African-American	30 (10.4%)	60 (14.4%)	
Hispanic, Race Specified	30 (9.4%)	40 (9.8%)	
Hispanic, No Race Specified	30 (11.1%)	50 (11.6%)	
Asian, Non-Hispanic	10 (2.4%)	10 (2.3%)	
Native Hawaiian or Other Pacific Islander	<10 (1.0%)	10 (2.3%)	
Native American or Alaskan Native	10 (3.4%)	20 (4.7%)	
More Than One Race	10 (2.0%)	10 (2.6%)	

Note. All values of n rounded to nearest 10. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 8

Comparison of Continuous Variables for Included and Excluded RD Samples

Variables	Included Mean (*SD*)	Excluded Mean (*SD*)	Independent Samples *t*-test (*df*)
Kindergarten SES	-0.25 (*0.67*)	-0.38 (*0.68*)	2.53 (*680*)*
Age at Fall Assessment Kindergarten	69.26 (*5.02*)	69.98 (*5.96*)	-1.67 (*660*)
Reading Assessment IRT Scale Score			
K1	28.77 (*4.30*)	28.35 (*4.27*)	0.94 (*400*)
K2	35.67 (*6.22*)	34.44 (*6.14*)	2.03 (*440*)*
1st Grade	53.44 (*11.28*)	50.96 (*11.13*)	2.26 (*450*)*
3rd Grade	91.73 (*22.10*)	87.72 (*22.50*)	1.87 (*430*)
5th Grade	118.40 (*22.14*)	111.36 (*24.38*)	3.23 (*470*)**
8th Grade	141.80 (*29.21*)	131.06 (*29.69*)	3.72 (*440*)***

Note. K1 = fall of kindergarten; K2 = spring of kindergarten. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 9

Comparison of Categorical Variables for Included and Excluded LRA Samples

Characteristics	Included Sample N (%)	Excluded Sample N (%)	Chi-Square (*df*)
Total	780	190	
Child Gender			0.62 (*1*)
Male	410 (52.4%)	100 (55.6%)	
Female	370 (47.6%)	80 (44.4%)	
Child Race/Ethnicity			9.71 (*7*)
White/Non-Hispanic	170 (22.3%)	60 (30.8%)	
African American/African-American	170 (21.8%)	40 (22.2%)	
Hispanic, Race Specified	120 (15.4%)	20 (10.8%)	
Hispanic, No Race Specified	190 (24.8%)	40 (21.6%)	
Asian, Non-Hispanic	50 (5.8%)	10 (7.6%)	
Native Hawaiian or Other Pacific Islander	10 (1.5%)	<10 (1.1%)	
Native American or Alaskan Native	50 (6.9%)	10 (5.4%)	
More Than One Race	10 (1.4%)	<10 (0.5%)	

Note. All values of *n* rounded to nearest 10. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 10

Comparison of Continuous Variables for Included and Excluded LRA Samples

Variables	Included Mean (*SD*)	Excluded Mean (*SD*)	Independent Samples *t*-test (*df*)
Kindergarten SES	-0.57 (*0.68*)	-0.63 (*0.87*)	0.99 (*890*)
Age at Fall Kindergarten Assessment	67.21 (*4.44*)	68.61 (*5.28*)	-2.80 (*870*)**
Reading Assessment IRT Scale Score			
K1	28.30 (*4.01*)	28.20 (*4.71*)	0.18 (*610*)
K2	35.94 (*5.33*)	35.21 (*5.73*)	1.33 (*720*)
1st Grade	52.42 (*9.25*)	49.74 (*10.67)*	3.14 (*830)***
3rd Grade	89.47 (*12.92*)	90.00 (*14.91*)	-0.48 (*950*)
5th Grade	116.19 (*16.10*)	115.29 (*18.90*)	0.03 (*960*)
8th Grade	136.92 (*24.49*)	136.74 (*29.48*)	0.07 (*890*)

Note. K1 = fall of kindergarten; K2 = spring of kindergarten; * $p <. 05$; ** $p < .01$; *** $p < .001$

Final Sample for Current Study

Participants assigned the longitudinal sample weight C1_7FC0 were retained for the current study. The longitudinal sample weight was assigned to children who had direct assessment data across six rounds of data collection (fall kindergarten, spring kindergarten, first grade, third grade, fifth grade, and eighth grade). This sample weight best fits the analyses based on the available weights published in the ECLS-K restricted dataset. Using a sample weight allows for generalizability to the kindergarten class of 1998-1999.

After identifying participants with the appropriate longitudinal sample weight, a field management system (FMS) variable that identified participants as having received special education services and those who had not received services at the first, third, and fifth grade time points was used to create two dichotomous groups. In other words, using this variable helped to ensure that there was no overlap between the two groups. Then, once participants were identified as either having received special education services or not received special education services between first and fifth grade, participants were selected for the final sample based on whether they met the designated criteria as a student with RD or LRA. For the RD group, participants were included if two criteria were met in first through fifth grade: (1) the special education teacher indicated a primary special education eligibility of learning disability, and (2) the special education teacher reported that the student had a reading goal on his/her IEP. If students met these criteria at any point between first and fifth grade, then they were included in the RD group. For the LRA group, students were included if they had a weighted average T-score below the 25th percentile using the first, third, and fifth grade time points. All students included in the LRA group had direct reading assessment scores in first, third, and fifth grade.

The final unweighted sample consisted of 1,080 children. Mean child age at each time point is reported in Table 11. Children in the final unweighted sample were approximately 32% White, non-Hispanic, 19% African American or African-American, 13% Hispanic, race specified, 21% Hispanic, no race specified, 5% Asian, 1% Native Hawaiian or Other Pacific Islander, 6% Native American or Alaskan Native, and 2% Multi-racial. About 56% percent of the sample is male. See Table 12 for child demographic characteristics.

Exclusionary Criteria

Participants were required to have the longitudinal sample weight C1_7FC0; 460 participants (RD group: *n* = 290; LRA group: *n* = 170) were excluded due to the absence of a nonzero weight. Additionally, given the interest in reading achievement over time between those with and without special education classification, children with intellectual disability, neurological conditions (e.g., traumatic brain injury, cerebral palsy), and developmental disabilities were excluded from the sample as to not to confound the results. Exclusion was based on school report of a diagnosis of developmental delay, intellectual disability, autism spectrum disorder, visual or hearing impairment, physical impairment, or multiple impairment at any of points between kindergarten and eighth grade. In the RD group, 50 children met these criteria, and in the LRA group, 60 children met these criteria. In addition, students were excluded from the RD group if they were identified as having RD in kindergarten (*n* = 50), as this could likely have been an inaccurate diagnosis (see below). Lastly, students were excluded from the RD group if they were identified as having RD for the first time between fifth and eighth grade (*n* = <10). No other exclusionary criteria were applied.

Sampling Weights

Descriptive analyses on the final sample include both unweighted and weighted means. Unweighted descriptive statistics refers to the ECLS-K population, in which each case was counted equally. Weighted descriptive statistics reflects the entire population of children attending kindergarten in 1998 in the United States, accounting for children who were not sampled in the ECLS-K data. The weighted sample represents population estimates from 1994 U.S. Census. Full sample weights and nesting variables (based on the Taylor Series method for computing standard errors) were applied to the latent growth model analyses to adjust for differential selection probabilities, reduce non-response bias and standard error, and allow for results that are generalizable to the population.

Table 11

Mean Ages at Each Time Point for Children Included in the Final Sample (n = 1,080)

Data Collection Wave	Age of Study Child (months)	
	Mean	SD
K1	67.77	4.70
K2	74.06	4.78
1st Grade	86.21	4.77
3rd Grade	110.33	4.72
5th Grade	133.91	4.75
8th Grade	170.63	4.73

Note. K1 = fall of kindergarten; K2 = spring of kindergarten. All values of n rounded to nearest 10.

Table 12

Demographic Characteristics of Final Sample (*n* =1,080)

Characteristic	Sample N (%)	Weighted N (%)
Child Gender		
Male	600 (55.6%)	378,500 (59.7%)
Female	480 (44.4%)	255,571 (40.3%)
Child Race/Ethnicity		
White/Non-Hispanic	350 (32.8%)	199,421 (31.5%)
African American/African-American	200 (18.7%)	166,970 (26.3%)
Hispanic, Race Specified	150 (13.8%)	86,649 (13.7%)
Hispanic, No Race Specified	230 (21.0%)	133,847 (21.9%)
Asian	50 (4.8%)	11,885 (1.9%)
Native Hawaiian or Other Pacific Islander	20 (1.4%)	4,419 (0.7%)
Native American or Alaskan Native	60 (5.9%)	23,676 (3.7%)
More Than One Race	20 (1.6%)	5,909 (0.9%)
Group Status		
Reading Disability	300 (27.5%)	160,014 (25.2%)
Low Reading Achievement	780 (72.5%)	473,957 (74.8%)

Note. All values of sample *n* rounded to nearest 10.

Definitions of Groups

Reading Disability (RD) Group

Overall, a total of 300 children with RD were included in the current study. During data collection, field supervisors were given the name of all students who received special education services along with the name of their primary special education teacher. The special education teacher then reported the child's main disability category. It is probable that identification of RD was based on the definition of a SLD from IDEA (1990, 1997), as participants were identified prior to the most recent reauthorization of IDEA in 2004. Using this definition, it is likely that schools used a discrepancy paradigm to identify students with RD. However, it is possible that some schools had begun to use a RTI model to identify those with RD.

In this study, the child's special education teacher's report of a learning disability (LD) in first, third, or fifth grade was used to identify the RD group. However, direct information about the child's learning disability type (e.g., reading, math) was unavailable, so the type of LD was determined based on the special education teacher's report of the child's IEP goals. Because this study is focused on understanding trends in reading achievement, children in this group were required to have at least one reading goal on their IEP. If the child had a reading goal, it was assumed that he/she displayed difficulty in reading and were subsequently categorized as having RD.

Children included in the RD group were identified to receive special education services at any point between first and fifth grade, based on special education teacher report as described above. RD identification at any point before fifth grade was chosen for three primary reasons. First, research suggests that early identification of LD in preschool and early kindergarten is often unreliable and inaccurate (Rathvon, 2004; Scarborough, 1998). Very young children may struggle academically in school; however, there are a number of reasons, other than a LD, that may explain why these children have academic difficulties. For example, they may struggle academically as a result of lack of formal education and/or previous instruction in basic reading skills (Rathbun & Zhang, 2016). This includes both formal preschool educational experiences as well as reading experiences in the home. Furthermore, young children may struggle in school if they have difficulty adapting to the classroom environment, following classroom expectations, as well as maintaining motivation and appropriate behavior during universal reading screenings and individual assessments (Rimm-Kaufman, Pianta, & Cox, 2000). Next, it is likely that there is a wide range of variation in the rate of development and learning of basic academic skills, such that some children take longer to understand basic reading skills than it does for others, but do not have LD (Share, Jorm, MacLean, & Matthews, 1984). Lastly, it is possible that young children identified as having LD will in fact be identified as having a cognitive impairment later on in their education. As a result, those identified with a RD in kindergarten were not included in the final sample.

Second, it is especially uncommon for children to first be diagnosed with RD after elementary school (i.e., fifth grade), as most children who have RD are identified in the early elementary school years. Even researchers who study late-emerging RD typically study fourth- and fifth-grade students, not middle school students (e.g., Compton, Fuchs, Fuchs, Elleman, & Gilbert, 2008; Leach, Scarborough, & Rescorla, 2003; Lipka, Lesaux, & Siegel, 2006). As a result, children diagnosed with a RD for the first time between fifth and eighth grades (n = <10) were not included in the RD group.

Third, including children identified between first and fifth grade captures most children who received special education services for a RD during their education. In other words, children who were identified both in early elementary school grades and later elementary school grades are included in this group. The children in this group are presumed to have had persistent reading difficulties throughout elementary school, permitting them to continue to qualify for special education services. Thus, the RD group consists of children who displayed persistent reading difficulties, which is hallmark of the diagnosis, in early elementary school grades as well as those who may have displayed a milder or late-emerging RD in fourth or fifth grade. Given that this study seeks to compare children who received special education services for RD to those with who did not receive such services but demonstrated LRA, it is rational to examine students who were identified as a child with a RD at any point between first and fifth grade.

It is also likely that some children in the RD group did not receive special education services for the entirety of the study. In other words, some children in the RD group exited from special education services prior to the conclusion of the study; that is, 80 children (26.7%) who were diagnosed with RD between first and fifth grade no longer received special education services for any classification by eighth grade. Still, it is important to include them in the RD group, because this study aims to examine the reading achievement trajectories and trends between those who have received special education services for RD, regardless of the length of time services were received, and those who were not identified but have LRA. Despite receiving special education services for varying lengths of time, these students received academic support via special education for at least some amount of time due to reading problems. Excluding children who exited from special education services would remove important variance in the reading achievement scores of the RD group.

Low Reading Achievement (LRA) Group

Overall, a total of 780 non-identified children with LRA in third grade were included in the current study. Children were identified as having LRA between first and fifth grade using two criteria. First, children with LRA were those who did not receive special education services at any point before fifth grade (i.e., non-identified). Second, children had a weighted average on the direct reading assessment below the 25th percentile using the first, third, and fifth grade time points. This average score was a weighted average derived from the first-grade T-score weighted 0.25, the third-grade T-score weighted 0.5, and the fifth-grade T-score weighted 0.25. If the weighted average T-score was less than 44 (i.e. 25th percentile), then the child was included in the LRA group. The 25th percentile was chosen as the cutoff, because researchers who study students with LRA often define their samples as those who are performing below the 25th percentile on standardized measures (e.g., Fletcher et al., 1994; Francis et al., 1996; Shaywitz, Fletcher, Holahan, & Shaywitz, 1992; Shinn, Ysseldyke, Deno, & Tindal, 1986).

A weighted average of three time points was used to identify the LRA group in order to develop a more representative score of children's achievement, as opposed to making LRA identification based on one single test score. The third-grade test score was more heavily weighted than the first and fifth grade time points for two main reasons. First, from a developmental perspective, it is expected that by third grade, children have learned how to read. Research suggests that with practice children become fluent and accurate readers between first and third grade (Kuhn & Stall, 2003). From kindergarten to third grade, the major focus is on teaching foundational reading skills (Etmanskie, Partanen, & Siegel, 2016). By fourth grade, however, the focus shifts from "learning to read" to "reading to learn" (Chall, 1983). Consequently, third grade is the point at which children should have adequate reading skills, making it appropriate to weight heavier than the other two time points to allow for the examination of the reading assessment scores of those who were performing well.

Second, it is important that the LRA group be identified for the sample at a points at which the LRA group is similar to the RD group in terms of reading assessment scores in order to compare their growth trajectories between kindergarten and eighth grade. Table 13 displays a comparison test of the reading assessment T-scores between the two groups. T-scores, as opposed to IRT scale scores, were used for this comparison, because they are meaningful for interpretation. In first grade, the mean T-score was 39.43 for the RD group and 39.17 for the LRA group. In third grade, the mean T-score was 37.97 for the RD group and 37.87 for the LRA group. In fifth grade, the mean T-score was 39.94 for the RD group and 39.45 for the LRA group. The independent samples t-tests at each of these time points were not significant, suggesting that the two groups were performing at very similar levels in terms of their reading performance based on the mean T-scores between first and fifth grade.

Table 13

Comparison of Mean T-Scores between RD and LRA Groups

Data Collection Wave	RD T-Score Mean (*SD*)	LRA T-Score Mean (*SD*)	Independent Samples *t*-test (*df*)
K1	42.74 (*6.28*)	41.88 (*6.47*)	1.79 (*810*)
K2	40.74 (*7.70*)	41.23 (*6.58*)	-0.96 (*870*)
1st Grade	39.43 (*8.62*)	39.17 (*6.99*)	0.49 (8*50*)
3rd Grade	37.97 (*9.27*)	37.87 (*6.24*)	0.21 (*1040*)
5th Grade	39.94 (*7.83*)	39.45 (*5.56*)	1.13 (*1060*)
8th Grade	42.04 (*8.31*)	40.49 (*6.58*)	3.12 (*1050*)**

Note. K1 = fall of kindergarten; K2 = spring of kindergarten. * $p < .05$; ** $p < .01$; *** $p < .001$.

Missing Data

Longitudinal research studies often have large amounts missing data due to attrition (e.g., move school, decline participation, death), and consequently, missing data must be analyzed. In the current sample, when examining only the 10 variables that were included in the final model, 6.7% of values were missing, and 34.1% of cases had missing values. Eight variables (80%) were missing 0.3-24.8% of data. Please see Table 14 for a summary of missing data.

Generally, it is recommended that if the proportion of missing data is small (i.e., less than 5%), then traditional methods (e.g., listwise or casewise deletion; single imputation) of handling missing data may be appropriate (Graham, 2009; Schafer, 1999). However, researchers suggest that even when the percentage of missing data is small, traditional missing data methods may lead to serious bias (Graham, 2009; Schafer, 1999). Deleting cases with missing data assumes that the data are missing completely at random, which is unlikely in most real-world research settings. Furthermore, Graham (2009) states that more advanced missing data methods should be used even when less than 5% of data are missing. In the current study, while only 6.7% of values were missing, 34.1% of cases had missing values, and using traditional methods to delete cases with missing values would likely lead to bias. Additionally, three variables had greater than 5% of missing data. As a result, full information maximum likelihood (FIML) was used to account for missing data.

Full Information Maximum Likelihood (FIML)

Although a number of valid methods have been developed to handle missing data (e.g., multiple imputation [MI], expectation maximization [EM]), FIML was chosen to handle missing data in the current study. Research supports the equivalence of FIML, EM, and MI for handling missing data (Collins, Schafer, & Kam, 2001; Schafer & Graham, 2002; Graham, Olchowski, &

Gilreath 2007; Dong & Peng, 2013). Therefore, FIML is an appropriate technique to handle missing data in the study at present.

Unlike MI, FIML does not impute plausible values for the missing values; rather, it integrates incomplete values into the estimation routine (Grimm, Ram, & Estabrook, 2017). In FIML, the marginal probability of observing the measured variables for each individual is computed over the variables with missing data (Allison, 2012 as cited in Grimm et al., 2017). In other words, FIML identifies the pattern of missing data and models the observed variables in groups of individuals who have the same pattern of missing data (Grimm, Ram, & Estabrook, 2017).

Rationale for FIML. According to Grimm et al. (2017), FIML estimation has become the main approach to handle missing data in structural equation modeling (SEM), and most SEM programs, such as MPlus, use FIML as a default. In the current study, FIML estimation was chosen over MI for two specific reasons. First, in MPlus, more fit statistics are available when FIML is used to estimate missing data as compared to when MI is used to estimate missing data (Muthén & Muthén, 2012). Similarly, there is little research on proper procedures for pooling fit statistics across multiple datasets when MI is used (Muthén & Muthén, 2012).

Second, FIML estimation in MPlus was chosen because the software allows for the inclusion of auxiliary variables. The inclusion of axillary variables is important because in order to use FIML the data must be missing completely at random (MCAR) or missing at random (MAR). Data MCAR occur when missing values have no relationship with other known or unknown values in the data; however, the MCAR mechanism is rarely upheld, as the missingness is usually related to some other observed variable (Haukoos & Newgard, 2007). When the data can be explained by observed variables, but not by the missing values, the data are said to be

MAR (Haukoos & Newgard, 2007). In other words, when the data are MAR, missing values may be related to certain observed predictor variables, but not to the missing values. Accordingly, when the missing data are related to either unobserved predictor variables or the missing values themselves, the data is MNAR (Haukoos & Newgard, 2007). FIML works under the assumption of MCAR and MAR, but not MNAR.

To test for MCAR, Little's MCAR test was used. Results indicated that the MCAR mechanism was not upheld in the current sample (χ^2 (120) = 293.57, $p < .001$). Although the data were not MCAR, it is likely that that the data were MAR. Again, this means that the missing values were likely related to observed variables, rather than to unobserved variables or the missing values themselves. Although there is no statistical test to determine MAR versus MNAR (Acock, 2012), the reasons for missingness can be examined from a theoretical standpoint. There were three variables with missing data greater than 5% (i.e., Fall Kindergarten Reading IRT Scale Score; Spring Kindergarten Reading IRT Scale Score; First Grade Reading IRT Scale Score).

Three possible reasons may explain why the percentage of missing data is higher in these reading achievement variables at the first three data collection time points. First, children who attended schools that refused to participate in the fall kindergarten wave but then decided to participate in the spring kindergarten wave would have been missing data in the fall of kindergarten. Second, children who failed an English language screening test in either kindergarten or first grade did not take the reading assessment, and, therefore, would have had missing data at these early time points. Third, children who were unable to complete the assessment as a result of challenging behaviors may have had missing data at these time points. To account for these potential explanations of missing data, an inclusive imputation model, which utilizes many predictor, or auxiliary variables, was used with FIML estimation. It is important to note the MPlus only allows for continuous variables to be included as auxiliary variables with FIML (Muthén & Muthén, 2012). Please see Table 15 for a list of auxiliary variables included in the model.

Lastly, Table 16 contains the means before and after FIML estimation. Of note, the means changed minimally after estimation. Thus, it is likely that the FIML estimation produced plausible and appropriate results.

Table 14

Summary of Missing Data for Analysis Model Variables

Variable	Missing		Valid N
	N	Percent	
K1 Reading IRT Scale Score	270	24.8%	810
K2 Reading IRT Scale Score	200	18.9%	880
1st Grade Reading IRT Scale Score	120	11.4%	960
K1 Continuous SES	50	4.3%	1030
8th Grade Reading IRT Scale Score	40	3.3%	1040
3rd Grade Reading IRT Scale Score	30	3.0%	1050
5th Grade Reading IRT Scale Score	10	1.3%	1070
K1 Race	<10	0.3%	1080
K1 Age at Assessment	0	0.0%	1080
K1 Gender	0	0.0%	1080

Note. The variables are listed from highest percentage of missing data to lowest percentage of missing data. K1 = fall of kindergarten; K2 = spring of kindergarten. All values of *n* rounded to nearest 10.

Table 15

Auxiliary Variables Included in Model

Variable	Waves Included[A]
Continuous SES	K1-8th grade
Child age	K1-8th grade
Child's total OLDS score	K1-1st grade
Child's Spanish total OLDS score	K1-1st grade
Approaches to learning (parent rating scale)	K1-1st grade
Self-control (parent rating scale)	K1-1st grade
Social interaction (parent rating scale)	K1-1st grade
Sad/lonely (parent rating scale)	K1-1st grade
Impulsive/overactive (parent rating scale)	K1-1st grade
Approaches to learning (teacher rating scale)	K1-5th grade
Self-control (teacher rating scale)	K1-5th grade
Interpersonal (teacher rating scale)	K1-5th grade
Internalizing problems (teacher rating scale)	K1-5th grade
Externalizing problems (teacher rating scale)	K1-5th grade
Self-control and interpersonal combined (teacher rating scale)	3rd grade-5th grade
SDQ interest/competence in reading (child rating scale)	3rd grade-8th grade
SDQ interest/competence in math (child rating scale)	3rd grade-8th grade
SDQ interest/competence in all subjects (child rating scale)	3rd grade-5th grade
SDQ interest/competence in peer relations (child rating scale)	3rd grade-5th grade
SDQ externalizing problems (child rating scale)	3rd grade-5th grade
SDQ internalizing problems (child rating scale)	3rd grade-8th grade
Locus on control (child rating scale)	8th grade
Self-concept (child rating scale)	8th grade
General knowledge ARS score (teacher rating scale)	K1-1st grade
Literacy ARS score (teacher rating scale)	K1-5th grade
Math ARS score (teacher rating scale)	K1-8th grade
Science ARS score (teacher rating scale)	3rd grade-8th grade
Social studies ARS score (teacher rating scale)	3rd grade
Oral ARS score (teacher rating scale)	8th grade
Writing ARS score (teacher rating scale)	8th grade
General knowledge routing test score	K1-1st grade
General knowledge IRT scale score	K1-1st grade
General knowledge T-score	K1-1st grade
Highest reading proficiency level	K1-8th grade
Reading routing test score	K1-8th grade
Reading IRT scale score	K1-8th grade
Reading T-score	K1-8th grade
Highest math proficiency level	K1-8th grade
Math routing test score	K1-8th grade
Math IRT scale score	K1-8th grade
Math T-score	K1-8th grade
Science routing test score	3rd grade-8th grade

Table 15 (cont'd)

Science IRT scale score	3^{rd} grade-8^{th} grade
Science T-score	3^{rd} grade-8^{th} grade

Note. [A]A variable was included in the imputation model for each wave included in the range listed. For example, child age was included at every wave, while child's total OLDS score was only included in kindergarten and first grade. K1 = fall of kindergarten; K2 = spring of kindergarten.

Table 16

Continuous Variable Means Before FIML Estimation and After Estimation

	Entire Sample		RD Group		LRA Group	
Variables	Before Estimation Mean	After Estimation Mean	Before Estimation Mean	After Estimation Mean	Before Estimation Mean	After Estimation Mean
Kindergarten SES	-0.48	-0.49	-0.25	-0.26	-0.57	-0.57
Age at Fall Kindergarten Assessment	67.77	67.77	69.26	69.26	67.21	67.21
Reading Assessment IRT Scale Score						
K1	28.45	27.97	28.77	28.27	28.30	27.86
K2	35.86	35.52	35.67	35.14	35.94	35.65
1st	52.72	52.58	53.44	52.88	52.42	52.41
3rd	90.07	89.86	91.73	91.14	89.47	89.38
5th	116.78	116.74	118.40	118.31	116.19	116.15
8th	138.18	137.90	141.81	140.96	136.92	136.74

Note. K1 = fall of kindergarten; K2= spring of kindergarten.

Variables and Measures

A description of the reading achievement variable as well as the covariates of interest is provided below. Please see to Table 17 for a list of variables.

Reading Achievement

For the current study, reading achievement is defined as the child's performance on the direct cognitive assessment of reading skills at each wave of data collection (i.e., fall kindergarten, spring kindergarten, first grade, third grade, fifth grade, eighth grade). It is important to note that all children with disabilities (i.e., those with IEPs, IFSPs, or 504 Plan) were provided accommodations during the assessments at each wave. The section below describes the development of the ECLS-K direct reading assessment.

Development of the reading achievement measure. The development of the direct cognitive battery, which included the direct reading assessment, consisted of five steps. First, a background review was conducted on all available and current psychometric instruments and the constructs they intended to measure. Second, for domains and constructs deemed relevant for kindergarten and first grade, test specifications were developed. Third, an item pool was created. Fourth, the item pool was field tested to determine whether the items were statistically and psychometrically appropriate for carrying out the overall assessment goals. Lastly, the final test forms were compiled based on field test item statistics and specifications.

Development of ECLS-K reading achievement test specifications. Using the Reading Framework for National Assessment of Educational Progress (NAEP), the ECLS-K team developed the reading specifications. The 1992 and 1994 NAEP framework was used to develop ECLS-K test specifications for earlier grade-levels (i.e., kindergarten through fifth), and the 1992-2007 NAEP framework was used for the eighth-grade test specifications.

The ECLS-K team adapted the NAEP framework to create the ECLS-K test specifications. Because the NAEP framework began in fourth grade, the team decided to add additional skill categories important for assessing younger children. The final six specifications included: basic skills, vocabulary, initial understanding, developing interpretation, personal reflection, and critical stance. Table 18 displays the number of questions for each reading skill assessed at the different grade levels over time.

Development of item pool. Once the test specifications were in place, a team of writers from the Educational Testing Service (ETS), elementary school curriculum specialists, and teachers created an item pool. Test items were adapted or borrowed from a number of published tests, such as the Peabody Individual Achievement Test-Revised (PIAT-R), Peabody Picture Vocabulary Test-Revised (PPVT-R), the Primary Test of Cognitive Skills (PTCS), the Test of Early Reading Ability (TERA-2), The Test of Early Mathematics Ability (TEMA-2), and the Woodcock-Johnson Tests of Achievement-Revised (WJ-R). Around 200 items were created for each subject area of the kindergarten and first-grade direct cognitive assessment. Subsequent items pools were created in a similar manner for the third-, fifth-, and eighth-grade assessments. All items were reviewed based on the appropriateness of content, difficulty, and relevance as well as issues related to cultural sensitivity. Once items passed the review process, they were constructed into field test booklets.

Field testing. Field testing occurred for each wave of data collection. Kindergarten and first-grade field testing was conducted together, and the third- and fifth-grade field testing was conducted together. The eighth-grade field testing was conducted on its own. Field testing was used to assess the feasibility of the assessment, whether a vertical scale could be created, the viability of a two stage routing procedures, and the psychometric properties of the assessment.

Kindergarten and first-grade field testing utilized approximately 1,500 students in the fall of 1996 and 1,500 first graders in the spring of 1997. Preliminary field testing for the third- and fifth-grade direct cognitive assessments occurred in the spring of 1999. A small number of children (n = 50) participated in these field tests, and a large number of items were tested. In the spring of 2000, a full-scale field test was conducted on approximately 1,800 children, using revised questions from the preliminary field test. Lastly, for the eighth-grade direct cognitive assessment, a field test was conducted to evaluate the psychometric properties of test items in the spring of 2006, and in the fall of 2006, a pilot test was carried out to determine whether procedures worked smoothly and produced accurate tests scores.

For the kindergarten and first-grade item pools as well as the third- and fifth-grade item pools, construct validity was evaluated. Construct validity of the ECLS-K kindergarten and first-grade reading item pool was supported by its high correlation (i.e., mid- to upper-eighties) with the Kaufman Test of Early Achievement (i.e., KTEA). Construct validity of the ECLS-K third- and fifth-grade reading item pool was supported by its high correlation (i.e., low eighties) with the Woodcock-McGrew-Werder Mini-Battery of Achievement (MBA). The eighth-grade field testing did not utilize a validation instrument to demonstrate construct validity, because the design and content of the eighth-grade assessment was based on previously validated surveys. According to Najarian, Pollack, and Sorongon's (2009) psychometric report, high correlations that resulted from previous validations provided support for strong construct validity of the ECLS-K eighth-grade item pool. For all grade-level field tests, items were included in the pool for final test forms if they demonstrated acceptable item analysis and IRT parameters (see **item response theory (IRT)** section below), had no differential item functioning (DIF) issues related to subgroup membership (see **differential item functioning (DIF)** section below), and showed increases in the percent correct between grade levels.

Differential item functioning (DIF). Differential item functioning (DIF) was used to determine whether a subgroup (e.g., African American students) performed differently on certain items when matched on performance to a reference group (e.g., White students). DIF procedures identified items when there was unexpected differential performance. After the DIF procedures identified certain items with large effect sizes, independent reviewers examined the content of these items. Items that were found to have content or a presentation that could be problematic for a certain subgroup were dropped from the item pool. However, items that showed differential performance between groups a result of children having different skill levels, as opposed to being due to subgroup membership, were retained.

Development of test forms. After a final item pool was established for each grade-level assessment, test forms were developed on the basis of item quality and reliability, item difficulty, floor and ceiling effects, longitudinal score scale, curriculum relevance, framework specifications, and practical issues (e.g., administration time). Items were also chosen for the final forms based on whether they met specific IRT criteria (see **item response theory (IRT)** section below). There were a small number of test items that did not meet the designated criteria but were still included for other reasons, such as framework specifications, links to a selected reading passage, or overlap with previous grade-level assessments.

Development of longitudinal scale. To allow for accurate measurement of abilities across time, ECLS-K put each grade-levels' assessment scores on a common scale, which required items from one grade-level assessment to overlap with the subsequent grade-level assessment. There was a total of 81 common items shared amongst all grade-level assessments. Additionally, a small second-grade "bridge" sample was used to remediate the gap in ability levels between the first-grade and third-grade time points. That is, second-grade students were needed in order to place items reliably along the difficulty scale, so gains from first to third grade could be measured. Because the distributions overlapped between third and fifth grade as well as between fifth and eighth grade, a fourth-grade “bridge” sample and a sixth and seventh-grade “bridge” sample, respectively, were not needed. To create this continuous longitudinal scale and determine whether the overlapping items were functioning appropriately across the different test forms, a statistical methodology called item response theory (IRT) was used.

Item response theory (IRT). Item response theory (IRT) was used to create a scale to compare scores across the different grade levels. In order to calibrate this scale, the test items within each subject area were mostly unifactorial, meaning that the same factor was underlying all test forms. As mentioned above, certain test items overlapped across adjacent forms and some content areas were represented in all grade forms, which allowed for the creation of the IRT scale. Responses for each of the shared items were combined for all rounds, and a single set of item parameters were estimated for each.

The ECLS-K IRT was based on a three-parameter model, which included a parameter for guessing, difficulty, and discriminability. The model used the pattern of responses (i.e., right, wrong, omitted) as well as the difficulty, discriminating ability, and “guess ability” of each item, placing each respondent at a particular point, θ (theta), on a continuous ability scale. Taken together, the three parameters for each item define a logistic function that represents the probability of a child answering the item correctly. The parameters were also used to create the IRT scale scores, T-scores, and proficiency scores, which will be described below in the **Reading Achievement Scores** section**.**

After the parameters were estimated, actual performance on the shared items in each round was compared with performance predicted by the IRT item and ability parameters. This allowed for the identification of discrepancies in terms of item difficulty at the different grade-levels. If the items displayed large differences in relative difficulty over time, then they would not be considered common items to compute estimates in gains over time. For the ECLS-K assessment, all reading items showed small differences between the actual and predicted performance, indicating common functioning of the items over time and good fit to the IRT model. Accordingly, once the final scoring procedures were developed and validated, the direct reading assessment was ready for administration.

Administration of the Reading Achievement Measure

The sections below describe the process for the language screening and routing procedures as well as test administration at each wave of data collection used in the current study.

Language screening. In the fall of kindergarten, if children were identified from their school records or by their teacher as speaking a language other than English as their primary language, then they were administered a language screening assessment prior to the reading assessment. This language screening was used to determine whether a child would be able to understand and respond to assessment questions in English. Spanish-speaking children who did not pass this language screener, based on an established cut-off score, received a reduced and translated version of the ECLS-K assessment. This reduced and translated version did not include the direct reading assessment. Non-Spanish speaking children who did not pass the language screening were excluded from the fall kindergarten assessment, and those who passed the language screening received the full English direct assessment.

In the spring of kindergarten and in the spring of first grade, children who passed the language screening at the previous assessment were immediately routed to take the reading assessment in English. Those who did not pass the language screening at the previous assessment retook the language screening and were routed according to their language screening score in the same manner as in the fall of kindergarten. After first grade, the language screening was no longer administered, and all children, regardless of English language proficiency, were administered the full English direct reading assessment. In sum, in the fall and spring of kindergarten and in first grade, children who were not proficient in English, based on their language screening score, did not take the direct reading assessment.

Two-stage routing procedures. In order to increase measurement accuracy and decrease administration time, the direct cognitive assessment occurred in two stages at all waves of assessment. The first stage was a routing stage in which 12 to 20 reading items ranging in level of difficulty were administered to all children. A child's performance on the routing stage determined what form was used at the second stage of testing.

For the kindergarten through fifth-grade reading assessments, there were low, middle, and high difficulty second-stage forms. In eighth grade, the reading assessment was changed to only include two second-stage forms: a low and a high form. Accordingly, the routing stage provided an estimate of the child's achievement level, so that the form with appropriate level of difficulty could be administered during the second stage. In the second stage, children answered questions that were appropriate for their level of ability. Therefore, frustration from answering too easy and/or too hard of questions along with administration time was minimized.

Developmental progression of reading achievement. Based on Chall's (1983) stages of reading development, learning to read is a cumulative process. In other words, once basic reading skills are attained, more complex understanding and analysis of written text can occur. This developmental and cumulative process of reading is represented in the kindergarten through eighth-grade reading assessments administered to participants in the current study. As students progressed from kindergarten through eighth grade, basic reading skills became less of a focus in the assessment, while more complex comprehension became more of a focus. The different proficiency levels assessed at each level, as described below, further illustrate this developmental progression.

Kindergarten reading achievement. Reading achievement scores from both the fall of kindergarten (1998) and spring of kindergarten (1999) are used for this study. Testing at both of these time points was conducted in a one-on-one testing session via CAI. The fall testing session took approximately 50-70 minutes beginning in September and ending in December, and it included cognitive, psychomotor, and physical components. In the spring, the assessment included only the cognitive and physical components and was conducted between March and June. At both time points, the reading assessment, which was included as part of the cognitive assessment battery consisted of questions in three subject areas: language and literacy (i.e., reading), mathematical thinking, and general knowledge. Children responded verbally or by pointing, and they were not asked to write or explain their reasoning.

The language and literacy, or reading assessment, in both the fall and spring of kindergarten were identical. It included questions designed to measure basic skills (i.e., print familiarity, letter recognition, beginning and ending sounds, rhyming sounds, word recognition), vocabulary (i.e., receptive vocabulary), and comprehension (i.e., listening comprehension, words in context). Five proficiency levels were evaluated during the reading assessment to reflect a sequence of skill development. The five levels were: (1) identifying upper- and lower-case letters of the alphabet by name (letter recognition); (2) associating letters with sounds at the beginning of words (beginning sounds); (3) associating letters with sounds at the end of words (ending sounds); (4) recognizing common words by sight (sight words); and (5) reading words in context (comprehension of words in context).

First-grade reading achievement. First-grade reading achievement scores from the spring of 2000 are used for this study. Testing was conducted in a one-on-one testing session via CAI. The direct child assessment took approximately one hour, and it included cognitive and physical components. Like the kindergarten direct cognitive assessment, the first-grade direct cognitive assessment focused on three subject areas: language and literacy (i.e., reading), mathematical thinking, and general knowledge. Children responded verbally or by pointing, as they did in the kindergarten assessment. The content of the first-grade reading assessment was identical to the kindergarten reading assessment, except that the number of reading items was increased by adding more difficult vocabulary and text. Still, the same five proficiency levels from the kindergarten assessment were of focus during the first-grade assessment.

Third-grade reading achievement. Third-grade reading achievement scores from the spring of 2002 are used for this study. Testing was conducted in a one-on-one testing session via CAI. The direct child assessment took approximately 90 minutes, and it included cognitive and physical components as well as a newly added child-completed socio-emotional component. The third-grade cognitive assessment included reading and mathematics as were included in kindergarten and first grade, but the general knowledge domain was changed to a science domain. Also in third grade, children gave not only pointing and verbal responses but also responded to reading passages and questions in booklet format. The administrator read all questions to the child, but the child read the multiple-choice response options individually.

The third-grade reading assessment included items that were designed to measure phonemic awareness, single word decoding, vocabulary, and comprehension. Items from the two highest proficiency levels on the kindergarten and first-grade reading assessment were retained, and three additional proficiency levels were added. Thus, the third-grade proficiency levels included two from the kindergarten and first-grade reading assessment along with three others: (4) recognizing common words by sight (sight words); (5) reading words in context (comprehension of words in context); (6) making inferences using cues that were directly stated with key words in text (literal inference); (7) identifying clues used to make inferences (extrapolation); and (8) demonstrating understanding of author's craft and making connections between a problem in the narrative and similar life problems (evaluation).

Fifth-grade reading achievement. Fifth-grade reading achievement scores from the spring of 2004 are used for this study. Testing was conducted in a one-on-one testing session via CAI. On average, fifth-grade assessments were completed within 96 minutes, and it included the same components as the third-grade assessment (i.e., cognitive, physical, and socio-emotional) as well as a child-completed food consumption questionnaire. The direct cognitive assessment also evaluated the same academic areas as the third-grade assessment: reading, mathematics, and science. To complete the fifth-grade assessment, children answered with pointing and verbal responses and were also given reading passages and questions in booklet format. The administrator read all questions to the child, but the child read the multiple-choice response options individually.

The fifth-grade reading assessment retained the four highest proficiency levels from the third-grade reading assessment and added one higher proficiency level. Therefore, the proficiency levels for the fifth-grade reading assessment included: (5) reading words in context (comprehension of words in context); (6) making inferences using cues that were directly stated with key words in text (literal inference); (7) identifying clues used to make inferences (extrapolation); (8) demonstrating understanding of author's craft and making connections between a problem in the narrative and similar life problems (evaluation); and (9) comprehension of biographical and expository text (evaluating non-fiction).

Eighth-grade reading achievement. Eighth-grade reading achievement scores from the spring of 2004 are used for this study. Unlike the previous rounds of assessment, the eighth-grade assessments were administered in a group format using pencil and paper tests. On average, eighth-grade assessments were completed within 80 minutes, and it included the same components as the third- and fifth-grade assessments (i.e., cognitive, physical, and socio-emotional) as well as a child-completed food consumption questionnaire. The direct cognitive assessment also evaluated the same academic areas as the third- and fifth-grade assessments: reading, mathematics, and science. To complete the assessment, children were given a first stage routing booklet for all three subjects, and they responded on answer sheets. The administrator then scored the routing form responses, while the children completed the socio-emotional self-questionnaire. Then, based on their scores on the first stage routing tests, children were given second-stage booklets, where they responded directly in the booklets.

The eighth-grade reading assessment retained the lowest proficiency level as well as the two highest proficiency levels from the fifth-grade reading assessment and added one higher proficiency level. Therefore, the proficiency levels for the eighth-grade reading assessment included: and (5) reading words in context (comprehension of words in context); (8) demonstrating understanding of author's craft and making connections between a problem in the narrative and similar life problems (evaluation); (9) comprehension of biographical and expository text (evaluating non-fiction); and (10) evaluating complex syntax and understanding high-level vocabulary (evaluating complex syntax).

Reading Achievement Scores

There are a number of different types of scores available from the direct cognitive assessment of the ECLS-K. First, number-right scores are raw scores, consisting of the number of items a child answered correctly. Similarly, there are item cluster scores, which are raw scores for the number of correctly answered items on a small subset of items. Third, proficiency level scores reflect a child's specific skills within a content area. Two types of scores are reported for proficiency levels: one score is a single indicator of the highest level mastered, and the other score is a set of IRT-based probability scores for each proficiency level. Scores in the data file that represent the highest proficiency level mastered are the numerical value reflective of that proficiency level. A score of zero indicates non-mastery of the lowest proficiency level. The second type of proficiency score, proficiency probability scores, estimates the probability of mastery of each level, ranging from zero to one.

Of interest in the current study are IRT scale scores and T-scores. As mentioned above, T-scores are used to define the LRA group and IRT scale scores are used as the measure of reading achievement in the analysis model. To create the IRT scale scores and T-scores, IRT item parameters and ability estimates were used. For each item, a child's probability of a correct answer was computed, even for items that were not administered to the child. This probability of answering an item correct was calculated using the child's ability estimate and item parameters. Then, the probabilities for each item were added together to get a scale score for each domain, and this score represents an estimate of the number of items the child would have answered correctly if he/she had taken all 212 reading items. Therefore, the IRT scale score for each grade-level are comparable to one another, because it is based on the same number total number of items (i.e., 212). IRT reading scale scores range from 0 to 212. The weighted means and standard deviations of the reading IRT scale scores for the entire ECLS-K sample are presented in Table 19. Reliability of IRT reading scale scores ranges from 0.87 to 0.96.

Furthermore, the T-scores are theta estimates transformed to a metric of a mean of 50 and a standard deviation of 10. T-scores are an indicator of the child's performance compared to the population average and how this changes over time. The T-scores provide a standardized metric to determine which children in the sample fall at, above, or below a certain percentile. Reading T-scores range from 0 to 96. Reliability coefficients for T-scores are the same as the IRT scale scores, because the T-scores are derived from the IRT scale scores.

Covariates

The covariates described below are time invariant (i.e., do not change over time). These variables were chosen as covariates in order to minimize their differential effects on reading achievement over time. In addition to the covariates described in the literature review, grade retention is not be used a control variable, as it is important to examine how the variance of those who were retained contributes to reading achievement over time.

Gender. The gender covariate is a composite of the child's gender based on information from the fall kindergarten parent interview.

Race/ethnicity. The race covariate is a composite of the child's race based on parent report during the fall kindergarten interview. This composite included eight categories: White, non-Hispanic; African American or African-American, non-Hispanic; Hispanic, race specified; Hispanic, no race specified; Asian; Native Hawaiian or other Pacific Islander; Native American or Alaskan Native; and more than one race specified, non-Hispanic.

Socioeconomic Status (SES). The SES covariate used in the current study is a composite of three indicators (i.e., household income, parent education, and parent occupation), which is based on parent report of these variables during both the fall and spring kindergarten interview. Kindergarten SES was chosen as a time invariant covariate, because the current study is interested in the early effects of SES on neurological development and reading achievement. Further, SES only changes by about one tenth of a point on the continuous scale from kindergarten to eighth grade (see Table 20). It is also highly correlated with the SES variable at each other time point (see Table 21). These correlations range from 0.74 to 0.92. As a result, it is appropriate to control for SES using the kindergarten time point.

Because both parents did not always respond to all questions related to SES in both the fall and spring of kindergarten, there was frequent missing data for each of the indicators. This missing data ranged from 2% to 28% for the different indicators. Therefore, ECLS-K researchers used a hot deck imputation method to impute missing values of all SES components. Using this methodology, the value reported by one respondent was used for a similar person with missing data. Chi-squared Automatic Interaction Detector (CHAID), a categorical search algorithm, was used to determine the best predictors for assigning values to those with missing data. Moreover, when households only had one parent, the available indicators were averaged to compute SES.

SES is a continuous composite that ranges from -4.75 to 2.75, with higher SES corresponding to higher values. Each indicator was standardized to have a mean of 0 and a standard deviation 1. Further information about each of the indicators is discussed below.

Household income. Household income represents the primary caregiver's annual reported income. These data were collected from the spring kindergarten parent interview.

Parent education. During the fall kindergarten parent interview, the responding parent reported the highest level of education attained. Education level was recoded categorically from 1 to 9 (i.e., 1 = 8th grade or below; 2 = 9th to 12th grade, 3 = high school diploma/equivalent, 4= vocational/technical program, 5 = some college, 6 = Bachelor's degree, 7 = graduate/professional school/no degree, 8 = Master's degree, 9 = Doctorate or professional degree).

Parent occupation. Parents' occupation was recoded to reflect the average of the 1989 General Social Survey (GSS; Nakao & Treas, 1992). This was carried out by averaging the corresponding prestige score from the 1998 Census occupational category codes, which were previously matched to the ECLS-K occupation codes. Information about parent's occupations was collected in the fall of kindergarten

Age at fall kindergarten assessment. Age at fall kindergarten assessment was chosen as a time invariant covariate, because it is highly correlated with the child's age at assessment at each other time point (see Table 22). These correlations range from .94 to .99, suggesting that these variables are collinear. As a result, it is appropriate to control for age using the fall kindergarten time point.

The age covariate is a composite of the child's age in months taken at the fall assessment. It was calculated by determining the number of days between the child's assessment date and the child's date of birth and dividing by 30.

Table 17

Summary of ECLS-K Variables Corresponding to Constructs and Covariates

Construct	ECLS-K Variables	ECLS-K Data Source
1. Reading Achievement Fall Kindergarten	a. Reading Scaled Score (Fall Kindergarten)	a. Direct Child Cognitive Assessment
2. Reading Achievement Spring Kindergarten	a. Reading Scaled Score (Spring Kindergarten)	a. Direct Child Cognitive Assessment
3. Reading Achievement First Grade	a. Reading Scaled Score (First Grade)	a. Direct Child Cognitive Assessment
4. Reading Achievement Third Grade	a. Reading Scaled Score (Third Grade)	a. Direct Child Cognitive Assessment
5. Reading Achievement Fifth Grade	a. Reading Scaled Score (Fifth Grade)	a. Direct Child Cognitive Assessment
6. Reading Achievement Eighth Grade	a. Reading Scaled Score (Eighth Grade)	a. Direct Child Cognitive Assessment
7. Covariates	a. Gender b. Race c. SES d. Age (at kindergarten fall assessment)	a./b./c./d. Parent Interview Questionnaire

Table 18

Percentage of Test Items Representative of Each Reading Skill by Grade of Assessment

				Reading Comprehension Skills		
Grade Levels	Total	Basic Skills	Vocabulary	Initial Understanding/ Developing Interpretation	Personal Reflection	Critical Stance
Percent of Testing Time						
Kindergarten	100	40	10	35	10	5
1st Grade	100	40	10	35	10	5
Percent of Test Items						
3rd Grade	100	15	10	45	15	15
5th Grade	100	10	10	45	15	20
8th Grade	100	0	0	55	15	30

Table 19

Weighted Means and SDs of Reading IRT Scale Scores for Entire ECLS-K Sample

Variable	Weighted Mean	SD
K1 reading IRT scale score	35.47	9.86
K2 reading IRT scale score	46.52	13.88
1st grade reading IRT scale score	77.07	23.70
3rd grade reading IRT scale score	125.70	28.57
5th grade reading IRT scale score	148.67	26.85
8th grade reading IRT scale score	167.24	28.03

Table 20

Weighted Mean SES and SD of Final Sample Over Time

Variable	Weighted Mean	SD
Kindergarten SES[a]	-0.51	0.64
1st grade SES	-0.58	0.61
3rd grade SES	-0.59	0.65
5th grade SES	-0.60	0.65
8th grade SES	-0.64	0.65

Note. [a]SES for kindergarten is based on both fall and spring data.

Table 21

Correlation Matrix for SES

	Kindergarten SES	1st grade SES	3rd grade SES	5th grade SES	8th grade SES
Kindergarten SES	--	--	--	--	--
1st grade SES	.81***	--	--	--	--
3rd grade SES	.80***	.88***	--	--	--
5th grade SES	.74***	.85***	.92***	--	--
8th grade SES	.77***	.85***	.90***	.91***	--

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 22

Correlation Matrix for Age

	K1 Age	K2 Age	1st grade Age	3rd grade Age	5th grade Age	8th grade Age
K1 Age	--	--	--	--	--	--
K2 Age	.95***	--	--	--	--	--
1st grade Age	.95***	.99***	--	--	--	--
3rd grade Age	.98***	.94***	.95***	--	--	--
5th grade Age	.97***	.94***	.94***	.98***	--	--
8th grade Age	.98***	.94***	.95***	.98***	.98***	--

Note. Note. K1 = fall of kindergarten; K2 = spring of kindergarten.* $p < .05$; ** $p < .01$; *** $p < .001$.

Research Questions and Hypotheses

Table 23 contains a summary of the four research questions and their corresponding statistical analyses.

Question 1

To what extent do SES, gender, race, and age predict whether a student will be identified as having either RD or LRA between first and fifth grade?

Hypothesis 1a. SES will predict whether a student is identified with RD. Research indicates that children with lower SES are more likely to be found eligible for special education services for RD than are those who are not experiencing low SES. Accordingly, it is hypothesized that those with lower SES are more likely to be identified to receive special education services in school.

Hypothesis 1b. Gender will predict whether a student is identified with RD. The extant literature indicates that boys are more likely to be referred for special education services. As a result, gender is hypothesized to significantly predict RD special education eligibility.

Hypothesis 1c. Race will predict whether a student is identified with RD. Previous research suggests that children from racial minorities are disproportionally over-represented in special education, and, therefore, race is hypothesized to predict identification as a student with a disability.

Hypothesis 1d. Age at kindergarten assessment will predict whether a student is identified with RD. Younger children have been found to be more likely to receive a RD diagnosis, so it is hypothesized that younger children will be more likely to receive special education services.

Question 2

How does initial reading achievement in the fall of kindergarten compare between children who received special education services and non-identified children with LRA with SES, gender, race, and age controlled?

Hypothesis 2. No research has compared RD and LRA groups at the start of kindergarten, so how they will differ is unknown. Still, it is hypothesized that children with RD will have lower initial reading achievement scores in the fall of kindergarten compared with children with LRA. Research suggests that those with RD have the severest reading difficulties, as a result of phonological deficits. Prior to the fall of kindergarten, it is unlikely that participants received significant instruction in phonological awareness and the alphabetic principle, so children with RD will likely have the lowest achievement at this point.

Question 3

How does the pattern, or shape, of reading growth from kindergarten through eighth grade compare between children who received special education services and non-identified children with LRA with SES, gender, race, and age controlled?

Hypothesis 3. It is hypothesized that the two groups will have similar patterns, or shapes of reading growth, but that there will be differences in their growth rates. Based on the compensatory trajectory of reading development, it is hypothesized that children who receive special education services for RD will have a faster rate of growth than children with LRA who do not receive special education services. However, research that examined the compensatory trajectory only compared children with RD to those without, so this will be one of the first studies to compare the rates of change between those with RD and those with LRA over an extended period of time.

Question 4

How do reading achievement outcomes in the spring of eighth grade compare between children who received special education services and non-identified children with LRA with SES, gender, race, and age controlled?

Hypothesis 4. Minimal research has examined outcomes between the two groups later in their academic careers, so it is unclear how they will differ at the end of eighth grade. Nonetheless, given the research that suggests that children with RD have poorer academic outcomes than children without RD, it is hypothesized that the children in the RD group will have significantly poorer academic achievement at the end of eighth grade compared to students with LRA. This hypothesis is also based on research that suggests that those requiring special education services are those with the most difficult to remediate reading problems.

Table 23

Summary of Study Hypotheses and Analyses

Hypothesis	Analysis	Variables	Method
Hypotheses 1a-1d	Effect of covariates on group status	SES (independent predictor) Gender (independent predictor) Race (independent predictor) Age (independent predictor) Group status (RD vs. LRA; dependent)	Multiple logistic regression
Hypothesis 2	Intercept (initial level) analysis with kindergarten	SES (independent predictor) Gender (independent predictor) Race (independent predictor) Age (independent predictor) K-8^{th} grade reading achievement (dependent)	Multiple group latent growth model
Hypothesis 3	Shape analysis	SES (independent predictor) Gender (independent predictor) Race (independent predictor) Age (independent predictor) K-8^{th} grade reading achievement (dependent)	Multiple group latent growth model
Hypothesis 4	Intercept (initial level) analysis with eighth grade	SES (independent predictor) Gender (independent predictor) Race (independent predictor) Age (independent predictor) K-8^{th} grade reading achievement (dependent)	Multiple group latent growth model

Data Analyses

Statistical Package for the Social Science (SPSS) version 24.0 was used to prepare data and run preliminary analyses, including descriptive analyses. MPlus student version 7.4 (Muthén & Muthén, 2012) was used to model the data. To select the sample for the current study based on inclusionary and exclusionary criteria and run preliminary analyses, variables of interest were imported into SPSS. Final analyses were conducted with and without replicate weights. All data analyses were conducted on a freestanding computer in Michigan State University's Erickson Hall, as per the restricted-use license agreement. Approval for use of the ECLS-K restricted-use data set was obtained from the Institute of Education Sciences (IES) and Michigan State University Institutional Review Board (IRB; see Appendix).

Preliminary Analyses

Before testing the latent growth curve model, preliminary analyses were conducted. Descriptive statistics were run using variables of interest to provide a better understanding of the ECLS-K sample and constructs of interest. Descriptive statistics of the covariates and outcome variables are presented in Table 24. Chi-square tests were conducted to determine if there were significant differences on all categorical study variables between the RD and LRA groups (see Table 25 and Table 26). Independent samples t-tests were conducted to determine if there were significant differences on all continuous variables between the RD and LRA groups (see Table 27 and Table 28). Additionally, a correlation matrix (see Table 29 and Table 30) for all variables of interest were run to screen for multicollinearity, and the distribution of variables were examined using graphs to address potential problems with outliers and a skewed distribution.

Logistic Regression

Logistic regression was used to answer research question one. Logistic regression is the appropriate regression technique to use when the dependent variable is binary (Field, 2013). Thus, for research question one, the goal was to examine whether the covariates predict the likelihood of being identified as a member of the RD group or the LRA group. The covariates included as predictors in the logistic regression model were gender, race, SES, and age at fall kindergarten assessment. If the variables significantly predicted group membership, then they were to be included in the final latent growth model analysis.

Latent Growth Model (LGM) Analyses

Latent growth model (LGM) analyses was used to answer research questions two through six. LGM is a special case of SEM. It is considered one of the most powerful approaches to analyzing growth trajectories, because it allows for the examination of both within- and between-person differences in change over time. To use LGM, there are three major requirements: (1) a continuous dependent variable measured on at least three different occasions; (2) scores that have the same units across time; and (3) cases that are all tested at the same intervals, though the intervals do not need to be equal (Kline, 2011).

There are a number of advantages of using LGM over other methods, such as analysis of variance (ANOVA) and multilevel modeling (e.g., hierarchical linear modeling (HLM)). LGM allows for the investigation of interindividual differences in change over time, and the researcher can investigate antecedents and consequences of change (Preacher, Wichman, MacCallum, & Briggs, 2008). LGM provides group-level statistics (e.g., mean growth rate and mean intercept), tests hypotheses about specific trajectories, and allows for the incorporation of both time-varying and time invariant covariates (Preacher et al., 2008). LGM offers the same advantages of SEM, which includes the ability to evaluate the adequacy of models using fit indices and model selection criteria, the ability to account for measurement error by using latent repeated measures, and the ability to deal effectively with missing data (Preacher et al., 2008).

Despite these advantages, certain assumptions for LGM with maximum likelihood (ML) must be met. Most of the assumptions involve the distributions of latent variables (i.e., intercept, slopes, and disturbances), but because these variables are unobservable, their characteristics are assumed. It is assumed that (1) the means of all latent variables, error terms, and factors have zero variance; (b) the variances of all latent variables have zero means; (c) the means and variance of latent variables do not covary; and (d) the error variances do not covary with each other or with any variables except the measured variables they directly affect (Duncan, Duncan & Strycker, 2013). Lastly, to use ML estimation, it is assumed that the observed variables are derived from population distributions with multivariate normality and kurtosis (Preacher et al., 2008). Because the current sample does not uphold the multivariate normality assumption, maximum likelihood estimation with robust standard errors (MLR) was used to estimate model parameters. MLR parameter estimates are robust to non-normality and non-independence of observations (Muthén & Muthén, 2012).

In a typical linear application of LGM, six parameters are estimated. These include the mean intercept and slope, the intercept and slope variances and covariance, and a disturbance variance that remains constant over repeated measures (Preacher et al., 2008). The remaining parameters are typically constrained to zero or to values that are consistent with a certain change pattern (Preacher et al., 2008). In a linear model, the intercept and slope are the two latent variables of interest. These two latent variables, or factors, are not psychological constructs, but represent patterns of change in *y*. The intercept factor represents the level of the outcome measure, y, at which the time variable equals zero, and the slope factor represents the linear rate at which the outcome measure changes (Preacher et al., 2008). Loadings on the intercept factor are fixed to one to represent the influence of a constant on the repeated measures. Loadings on the slope factor are fixed to a linear progression to represent linearly increasing growth over time.

When growth does not appear to be linear, a non-linear LGM model can be used. One example of a non-linear LGM is called a latent basis growth model (McArdle & Epstein, 1987; Meredith & Tisak, 1990). In a latent basis growth model, an initial level and a shape factor are estimated. The shape factor is the predicted amount of change in the outcome variable of interest between at least two time points that are fixed. For example, the first time point in the model would be fixed at zero and the last time point in the model would be fixed at one. The time points in between are then freely estimated. As in a linear model, the total amount of change is perfectly related to the rate of change, because there is only one change component; in contrast to the linear model, the rate of change is not constrained to be constant across time in the latent basis model (Grimm, Ram, & Hamagami, 2011). Furthermore, a latent basis model can capture many non-linear change patterns because it does not have a specific functional form, and it rescales time for optimal fit (Grimm et al., 2011).

General procedures. In order to fully examine the reading growth trajectories of children with RD and children with LRA, a number of LGM models were examined. First, unweighted and weighted unconditional models (i.e., models without covariates) were fit to the data. The purpose of these models was to determine the longitudinal change in reading achievement across the entire sample (i.e., both children with RD and children with LRA). Once the nature of the longitudinal change was established, then weighted and unweighted conditional models (i.e., model with covariates) were fit to the data. These conditional models were used to examine the relationship among the latent growth factors and the covariates across the entire sample (i.e., both children with RD and children with LRA).

With the nature of the longitudinal change established as well as the relationship between the relationship among the latent growth factors and the covariates examined, multiple group LGM was conducted. For the multiple group analyses, the unweighted and weighted conditional models were used, because the nature of the longitudinal change had already been established across the entire samples using the unconditional model. Prior to running the multiple group LGM, the conditional latent basis model with residual covariances/regressions were run separately for each group. Once fit was adequate in each group, then three different multiple group LGMs were conducted. The first multiple group LGM determined whether there were group differences in the initial level of reading achievement, the second multiple group LGM examined whether there were group differences in the shape, or pattern of change in reading achievement over time, and the third multiple group LGM assessed whether there were group differences in the reading achievement outcomes in eighth grade. For each of these three analyses, there was an unconstrained and a constrained model (Acock, 2005; Duncan et al., 2006). In the unconstrained model, none of the estimates were set to be equal across groups. In the constrained model, the parameter of interest was constrained to be equal across groups. In the models examining group differences in the initial level, the initial levels were constrained to be equal. In the models examining group differences in the shape, the basis coefficients were constrained to be equal. In the models examining group differences in the eighth-grade outcomes, the eighth-grade time point was used as the initial level, and then the initial levels were constrained to be equal. In this model, the kindergarten fall time point was fixed to -1; the kindergarten spring, first grade, third grade, and fifth grade time points were freely estimated; and the eighth-grade time point was fixed to zero. The zero point of the time scale determines the time at which to interpret the parameter estimates (Rogosa & Willett, 1985), so by fixing the eighth-grade time point to zero, it can be interpreted as the initial level.

Model fit. Absolute and relative fit indices were used to determine model fit. The chi-square statistic, an absolute fit index, was examined in all analyses. A significant chi-square value indicates that the model is not a good fit. A non-significant chi-square index ($p > .01$) indicates that the null hypotheses is not rejected, and the model does not significantly differ from the true population that generated the data (Duncan et al., 2006). However, the chi-square statistic is affected by sample size, which may lead to rejection of the null hypothesis even when there are differences between the observed and predicted covariances (Kline, 2011). Thus, additional fit indices that are not affected by large sample sizes were used. Comparative Fit Index (CFI) and Tucker-Lewis Fit Index (TLI) are relative fit indices, which both account for sample size (Kline, 2011). The CFI and TLI values greater than or equal to 0.90 indicate adequate fit (Bollen & Curran, 2006), although some suggest more stringent cutoff values of greater than or equal to 0.95 (Hu & Bentler, 1999). Finally, the Root Mean Square Error of Approximation (RMSEA) will be used as an indicator of absolute fit. The RMSEA provides an estimation of the model with optimal parameter estimates that fits the covariance matrix for the population. An RMSEA value less than or equal to 0.07 with an upper limit less than or equal to 0.08 is considered adequate model fit (Hooper, Coughlan, & Mullen, 2008). Hu and Bentler (1999) suggest a more stringent RMSEA cutoff of less than or equal to 0.06.

Additionally, when comparing the model fit of different models, the Satorra-Bentler scaled chi-square difference test was used. This chi-square differences test uses the normal chi-square statistic and divides it by a scaling correction to better estimate chi-square under non-normality (Muthén & Muthén, 2012). In MPlus, the Satorra-Bentler scaled chi-square difference test must be used when using MLR (Muthén & Muthén, 2012).

For multiple group LGM, a chi-square significance test was conducted by calculating the chi-square difference statistic (i.e. the Satorra-Bentler scaled chi-square difference test) between the constrained and unconstrained models. If model fit significantly worsened when the parameter of interest was constrained, then group differences might exist (Kline, 2011). In other words, if the chi-square statistic was significant, this suggests the model parameters were unequal across groups.

In addition to the Satorra-Bentler scaled chi-square difference test for the multiple group LGM, a Wald test was used to examine group differences in the initial level and the outcomes in eighth grade. A Wald test is very similar to the chi-square difference test; however, it only requires estimating one model, such that it tests whether two parameters are significantly different from each other. According to the UCLA: Statistical Consulting Group (2017), the chi-square difference test and the Wald test will become asymptotically equivalent to each other as the sample size gets larger. In smaller samples, there may be slight differences between the two tests, but the two usually lead to the same conclusions (UCLA: Statistical Consulting Group, 2017). A Wald test was not conducted to test whether there were differences in the shape, or basis coefficients, because MPlus does not allow for equality constraints to be placed in a growth statement. Therefore, the Wald test was used to compliment the Satorra-Bentler scaled chi-square difference test for evaluating group differences in the initial level and the eighth-grade outcomes, although results of the Wald test and the Satorra-Bentler scaled chi-square difference test were expected to be similar.

Sensitivity Analyses

Sensitivity analyses help researchers to demonstrate the robustness of their results and conclusions (Thabane et al., 2013). Sensitivity analyses test the effects of changing certain aspects of the study (e.g., method of analyses, cut-points used to define groups, outliers, method used to handle missing data) on the observed results (Thabane et al., 2013). If the results of the sensitivity analyses are consistent with those from the primary analyses and lead to similar conclusions, researchers can be more confident that the underlying factors had little or no influence on the primary results and conclusions (Thabane et al., 2013).

In the current study, it is important to use sensitivity analyses to determine if the results of the study would change if the groups had been defined differently. The groups that were used in all of the analyses and serve as the basis for comparison were (1) RD group identified between first and fifth grade, and (2) LRA group identified using weighted average of first, third, and fifth grade T-scores. In each sensitivity analysis, either the RD or the LRA group was replaced with a group defined using different criteria. In other words, the four sensitivity analyses always maintained one of the original RD or LRA groups, while the other group was replaced with a group defined using different criteria.

The first sensitivity analysis maintained the original RD group but utilized a LRA group that was defined using only one test score. The original LRA included students who had a weighted average T-score from the first, third, and fifth reading assessments that fell below the 25^{th} percentile. Thus, question becomes, would the results change if the LRA group was defined using only one reading assessment T-score below the 25^{th} percentile, specifically from the third-grade assessment? The third-grade assessment was chosen, because based on Chall's (1983) stage theory, students are expected to know how to read by third grade, and by fourth grade, the focus shifts from “learning to read” to “reading to learn.” Accordingly, the first sensitivity analysis included students in the LRA group if they performed below the 25^{th} percentile on the direct reading assessment in third grade.

The second sensitivity analysis maintained the original LRA group but included students who were identified to receive special education services due to RD at any point between kindergarten and fifth grade. In the original RD group, students identified as having a RD in kindergarten were excluded from the sample due to possible inaccurate classifications. Nonetheless, conducting a sensitivity analysis with those students identified as RD in kindergarten helped to determine whether the results might change when including those students.

The third and fourth sensitivity analyses also maintained the original LRA group but included students who were identified to receive special education services due to RD at any point between first grade and fourth grade and first grade and third grade, respectively. Given that the core deficit of RD is widely suggested to be related to phonological awareness and related early reading skills, it is unlikely that students identified after third or fourth grade have the same type of reading deficits that are considered in phonological-deficit RD. Furthermore, researchers who study late-emerging RD typically study students before fifth grade, not older middle school students (e.g., Compton, Fuchs, Fuchs, Elleman, & Gilbert, 2008; Leach et al., 2003; Lipka, Lesaux, & Siegel, 2006). Accordingly, some may argue that students in fourth and fifth grade should not be included in the RD group, as their reading deficits are different than those identified in earlier grades. As a result, the final two sensitivity analyses examined whether the results of the study change when the RD group included students identified between first grade and fourth grades and first and third grades.

CHAPTER 4
RESULTS

Preliminary Analyses

Means and standard deviations of the covariates and outcome variables for the study sample are displayed in Table 24. Regarding the unweighted sample statistics, the mean kindergarten SES was -0.48 on a scale that ranges from -4.75 to 2.75. The mean kindergarten SES of the included sample was significantly lower than the mean kindergarten SES of the excluded sample, as described above, suggesting that this group of poor readers comes from families with lower socioeconomic backgrounds. Children in the included sample had a mean age of five years and six months of age in the fall of kindergarten, which was significantly younger than those in the excluded sample. The reading assessment IRT scale scores of the sample increased each year, but the greatest increases occurred between the springs of first and third grade and the springs of third and fifth grade. The mean IRT scale scores of the included sample were significantly less than the mean IRT scale scores of the excluded sample, as described above. Furthermore, the full sample included about 10% more boys than girls, whereas the difference in the excluded sample was only about 1%, and approximately 67% of the sample were children from diverse racial/ethnic groups. In the excluded sample, about 44% of the sample were children from diverse racial/ethnic groups.

It was also necessary to compare the descriptive statistics of the unweighted sample to those of the weighted sample. In general, the unweighted and weighted samples were relatively similar (see Table 12 and Table 24). Specifically, regarding child race/ethnicity, the unweighted sample had about 8% less African American children, about 3% more Asian children, and about 2% more Native American or Alaskan Native children than the weighted sample. All other differences in race/ethnicity between the unweighted and weighted samples were approximately

1% or less. For gender, the unweighted sample had 4% less boys than the weighted sample. Lastly, age, SES, and the reading assessment IRT scale scores differed only minimally between the unweighted and weighted samples.

These minimal differences between the unweighted and weighted samples are important to keep in mind when reading further about the LGM analyses and subsequent results. Many of the LGM analyses using the weighted data produced non-positive definite covariance matrices (i.e., negative slope and intercept variances) and indicated that the models were not the best fit for the weighted data, likely resulting from extremely large variances in the reading assessment IRT scale scores. Thus, although the purpose of using weighted data would be to allow for the generalization of results to the national kindergarten class of 1998-1999 and mitigate issues related to non-response bias, weighted analyses were not always permissible in the current study. Nonetheless, given the informal comparisons between the unweighted and weighted samples, which suggest that the two samples were not entirely different in terms of demographic information and average reading IRT scale scores, the results using the unweighted data may generalize somewhat comparably to the national kindergarten class of 1998-1999.

Differences Between RD and LRA Groups

Weighted and unweighted group differences are displayed in Tables 25, 26, 27, and 28. With respect to unweighted group differences, the RD group had about 12% more boys than the LRA group, and the RD group was about 40% racially/ethnically diverse, whereas the LRA group was about 78% racially/ethnically diverse. These unweighted group differences in gender and race/ethnicity were statistically significant. The LRA group also had a significantly lower SES than the RD group, and the LRA group was about two months younger than the RD group at the fall of kindergarten, which was a statistically significant difference, although perhaps not a clinically meaningful difference. Based on the unweighted data, the only significant differences between the two groups in reading assessment IRT scale scores were at the third- and eighth-grade time points, in which the RD group outperformed the LRA group.

As with the unweighted data, the weighted RD group had about 12% more boys than the weighted LRA group. In terms of race/ethnicity, the weighted RD group was about 48% racially/ethnically diverse, whereas the weighted LRA group was about 75% racially/ethnically. That is, the RD group became 8% more racially/ethnically diverse when the weighted data was used, but the LRA group remained relatively the same. The group differences in gender and race/ethnicity were statistically significant, as they were in the unweighted samples. In addition, the weighted samples' SES and age remained quite similar to that of the unweighted sample, and the differences between the RD and LRA groups' SES and age with the weighted data were also statistically significant. Lastly, in regard to the reading assessment IRT scale scores with the weighted samples, the scores at each grade level remained within two to three points of the scores using the unweighted data; however, there were more significant differences between the RD and LRA groups' scores when the weighted data were used. More specifically, the RD group significantly outperformed the LRA group at the fall kindergarten, third-grade, fifth-grade, and eighth-grade time points, and the LRA group significantly outperformed the RD group in the fall of kindergarten. Differences in the two group's scores at the first-grade time point were not significant. Still, it is important to note that while some of the differences in the reading assessment IRT scale scores in the weighted groups were statistically significant, such differences may not be clinically significant. For example, the two groups differed by three-tenths of a point and two-tenths of a point at the fall of kindergarten and spring of kindergarten time points, respectively. Consequently, the unweighted and weighted samples appeared relatively similar even when examining differences between the RD and LRA groups, such that the LGM analyses with the unweighted data may not be extremely different from the LGM analyses with the weighted data.

Table 24

Descriptive Statistics for Full Sample

Variable	Mean	SD	Range	Weighted Mean	Weighted SD
Kindergarten SES	-0.48	0.69	-4.47-2.25	-0.51	0.64
Age at Fall Kindergarten Assessment	67.77	4.70	45.77-86.40	67.67	4.55
Reading Assessment IRT Scale Score					
K1	28.45	4.11	21.07-47.61	28.27	3.96
K2	35.86	5.62	23.40-52.05	35.65	5.48
1^{st}	52.72	9.89	25.56-105.03	52.30	9.68
3^{rd}	90.07	15.52	51.46-155.11	89.71	15.47
5^{th}	116.78	17.94	67.79-172.78	116.26	18.23
8^{th}	138.18	25.86	86.63-204.64	135.65	26.10

Table 25

Unweighted Descriptive Statistics Comparison Between RD and LRA Groups for Categorical Variables

Characteristics	RD Sample N (%)	LRA Sample N (%)	Chi-Square (*df*)
Total	300	780	
Child Gender			11.62 (<*10*)**
Male	190 (64.0%)	410 (52.4%)	
Female	110 (36.0%)	370 (47.6%)	
Child Race/Ethnicity			144.66 (*10*)***
White/Non-Hispanic	180 (60.3%)	170 (22.3%)	
African American/African-American	30 (10.4%)	170 (21.8%)	
Hispanic, Race Specified	30 (9.4%)	120 (15.4%)	
Hispanic, No Race Specified	30 (11.1%)	190 (24.8%)	
Asian, Non-Hispanic	10 (2.4%)	50 (5.8%)	
Native Hawaiian or Other Pacific Islander	<10 (1.0%)	10 (1.5%)	
Native American or Alaskan Native	10 (3.4%)	50 (6.9%)	
More Than One Race	10 (2.0%)	10 (1.4%)	

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 26

Weighted Descriptive Statistics Comparison Between RD and LRA Groups for Categorical Variables

Characteristics	RD Sample N (%)	LRA Sample N (%)	Chi-Square (*df*)
Total	160,014	473,957	
Child Gender			2241.43 (*1*)***
Male	103,565 (64.7%)	274,935 (58.0%)	
Female	56,449 (35.3%)	199,022 (42.0%)	
Child Race/Ethnicity			43425.18 (*7*)***
White/Non-Hispanic	83,261 (52.0%)	116,160 (24.6%)	
African American/African-American	27,779 (17.4%)	139,190 (29.4%)	
Hispanic, Race Specified	16,608 (10.4%)	70,041 (14.8%)	
Hispanic, No Race Specified	24,266 (15.2%)	109,581 (23.2%)	
Asian, Non-Hispanic	1,609 (1.0%)	10,276 (2.2%)	
Native Hawaiian or Other Pacific Islander	355 (0.2%)	4,064 (0.9%)	
Native American or Alaskan Native	3,982 (2.5%)	19,694 (4.2%)	
More Than One Race	2,254 (1.3%)	3,755 (0.8%)	

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 27

Unweighted Descriptive Statistics Comparison Between RD and LRA Groups for Continuous Variables

Variable	RD Group		LRA Group		Independent Samples *t*-test (*df*)
	N	Mean (*SD*)	N	Mean (*SD*)	
Kindergarten SES	290	-0.25 (*.66*)	740	-0.57 (*.68*)	6.85 (*1030)******
Age at Fall Kindergarten Assessment	300	69.26 (*5.02*)	780	67.21 (*4.44*)	6.53 (*1080)******
Reading Assessment IRT Scale Score					
K1	260	28.77 (*4.30*)	550	28.30 (*4.01*)	1.52 (*810)*
K2	270	35.67 (*6.22*)	610	35.94 (*5.33*)	-0.67 (*870*)
1st	280	53.44 (*11.28*)	680	52.42 (*9.25*)	1.45 (*950*
3rd	280	91.73 (*21.10*)	770	89.47 (*12.92*)	2.07 (*1040**
5th	290	118.40 (*22.14*)	780	116.19 (*16.10*)	1.78 (*1060)*
8th	270	141.81 (*29.21*)	780	136.92 (*24.49*)	2.68 (*1050)***

Note. K1 = fall of kindergarten; K2 = spring of kindergarten. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 28

Weighted Descriptive Statistics Comparison Between RD and LRA Groups for Continuous Variables

Variable	RD Group		LRA Group		Independent
	N	Mean (*SD*)	N	Mean (*SD*)	Samples *t*-test (*df*)
Kindergarten SES	154,302	-0.30 (*0.65*)	455,745	-0.58 (*0.62*)	154.08 (*610045)****
Age Fall Kindergarten at Assessment	160,014	68.82 (*5.00*)	473,957	67.28 (*4.32*)	117.97 (*633969)****
Reading Assessment IRT Scale Score					
K1	140,708	28.49 (*4.20)*	353,092	28.18 (*3.86*)	24.70 (*493797)****
K2	147,619	35.49 (*6.24*)	380,432	35.71 (*5.15*)	-13.01 (*528049)****
1st	152,610	52.34 (*11.21*)	416,819	55.29 (*9.06*)	1.77 (*569427)*
3rd	142,412	92.55 (*21.02*)	464,025	88.83 (*13.18*)	79.76 (*606435)****
5th	153,879	121.02 (*23.28*)	471,324	114.70 (*15.94*)	119.43 (*625201)****
8th	150,756	142.67 (*29.53*)	468,165	133.39 (*24.47*)	121.57 (*618919)****

Note. K1 = fall of kindergarten; K2 = spring of kindergarten. * $p < .05$; ** $p < .01$; *** $p < .001$.

Correlations and Sample Distribution

Correlations between all student variables were computed separately for children in the RD group and children in the LRA group to determine whether the strength of the coefficients was similar or different between groups. Table 29 displays a correlation matrix for both groups. No study variables were highly correlated (r >.85), indicating no potential problems with multicollinearity (Kline, 2011).

Regarding univariate normality, Q-Q plots and histograms revealed non-normal distributions of the dependent variables and the standardized residuals of the dependent variables (i.e., reading IRT scale scores at each time point), such that the data appeared to be positively skewed. The Shapiro-Wilk Test of Normality for the dependent variables and the standardized residuals of the dependent variables indicated non-normal distributions ($p < .001$), with the exceptions of the standardized residuals for the kindergarten spring reading IRT scale score (p=.15) and the fifth-grade reading IRT scale score ($p = .06$). For multivariate normality, Mardia's Multivariate Normality Test, Henze-Zirkler's Multivariate Normality Test, and Royston's Multivariate Normality Test indicated non-normal multivariate distributions as well ($p < .001$). Nonetheless, non-normal distributions were expected given that the sample only included children with reading problems. Outliers were included in the study because the research questions aimed to examine group differences among children with reading difficulties, so it was expected that children would fall in the extreme ends of the spectrum in reading achievement.

Table 29

Correlation Matrix for RD (Upper) and LRA (Lower) Groups

	1	2	3	4	5	6	7	8	9	10
1	--	-.01	.04	-.09	.03	.02	.13*	.06	.05	.06
2	.04	--	-.24***	-.09	-.20**	-.17**	-.24***	-.23***	-.35***	-.27***
3	-.08*	-.12**	--	-.05	.26***	.27***	.36***	.34***	.39***	.37***
4	-.06	.04	-.03	--	.34***	.22***	.21***	.09	.07	.01
5	.00	-.18***	-.17***	.24***	--	.62***	.48***	.30***	.29***	.14*
6	.10*	.08	-.16***	.23***	.56***	--	.55***	.38***	.39**	.23***
7	.10**	-.11**	-.08*	.10***	.38***	.52***	--	.55***	.50***	.27***
8	.05	-.06	.11**	-.07	.13**	.14**	.16***	--	.72***	.49***
9	.01	-.07	.15***	-.09*	.14**	.14**	.13**	.42***	--	.65***
10	.08*	-.04	.19***	-.09*	.15***	.14**	.04	.28***	.51***	--

Note. * $p < .05$; ** $p < .01$; *** $p < .001$. 1=gender, 2=race, 3=kindergarten SES, 4=fall kindergarten age, 5=fall kindergarten reading IRT score, 6=spring kindergarten reading IRT score, 7=1st grade reading IRT score, 8=3rd grade reading IRT score, 9=5th grade reading IRT score, 10=8th grade reading IRT score.

Research Question 1: Predicting Group Membership

The first research question examined whether the covariates of interest (i.e., SES, gender, race/ethnicity, and age) predicted whether a student would have been identified as having a RD or LRA between first and fifth grade. In this model, the reference group was the RD group (i.e., RD=1, LRA=0), and the categorical/ordinal predictor variables were coded into dummy variables, which resulted in seven final predictors: SES, female, African American, Hispanic, Asian, other diverse racial/ethnic group, and age. Unweighted and weighted results are summarized in Table 30.

Parameter estimates for the unweighted logit model are given in terms of log-odds, which allows for the conversion to an odds-ratio. An odds-ratio represents the change in the odds of a child being identified for the RD group that is associated with a one-unit change in a particular predictor variable. An odds-ratio greater than one reflects an increase in the likelihood of being identified as a member of the RD group, and an odd-ratio less than one reflects a decrease in the likelihood of being identified as a member of the RD group.

Unweighted results indicated that all of the independent variables significantly predicted membership in the RD group. Because many of the initial odds-ratios were less than one, the ratios were converted (i.e., 1/ e^{β}) to be greater than one, which then can be interpreted in light of the group with the higher odds. For example, the odds of being identified as a student with RD, as compared to a student with LRA, was 0.73 times as large for girls as for boys. In other words, boys had 1.38 times greater odds, or a 38% greater chance, of being identified for the RD group than did girls. Regarding race/ethnicity, African American, Hispanic, Asian, and other racially/ethnically diverse students had significantly less odds of being identified for the RD group than did White students. That is, African American students were 4.74 times more likely to be identified for the LRA group than for the RD group; Hispanic students were 3.91 times more likely to be identified for the LRA group than for the RD group; Asian students were 4.43 times more likely to be identified for the LRA group than for the RD group; and students from other diverse racial/ethnic groups were 3.70 times more likely to be identified for the LRA group than for the RD group. In addition, for each unit increase in SES, students were 1.45 times more likely to be identified for the RD group than for the LRA group. Lastly, for each unit increase in age at the fall kindergarten assessment, students were 1.09 times more likely to be identified for the RD group than the for LRA group.

To summarize the above logistic regression results, when race/ethnicity, SES, and age were controlled, boys were more likely to identified for the RD group than were girls. With gender, SES, and age controlled, non-White students were significantly more likely to be identified for the LRA group, while White children were significantly more likely to be identified for the RD group. With gender, race/ethnicity, and age controlled, children with higher SES were significantly more likely to be identified for the RD group, and with gender, race/ethnicity, and SES controlled, older students were significantly more likely to be identified for the RD group than for the LRA group.

When conducting a logistic regression using replicate weights in MPlus, logit models are unavailable. Therefore, a probit model was used. Logit and probit models differ in the way the function is defined, but estimates usually do not differ greatly. However, odds ratios cannot be computed from probit model estimates. Weighted results indicated that all of the independent variables significantly predicted membership in the RD group, except for the gender dummy variable. Although gender did not predict group membership, it was still used as a covariate in the LGM model, given that gender is believed to play a substantive role in longitudinal reading development.

Table 30

Summary of Logistic Regression of Covariates Predicting Membership in RD Group

Variables	Unweighted Model[A]			Weighted Model[B]		
	β	**SE β**	e^{β}	**β**	**SE β**	e^{β}
Female	-0.32*	0.16	0.73	-0.10	0.13	N/A
Race/Ethnicity						
African American	-1.56***	0.24	0.21	-0.60**	0.23	N/A
Hispanic	-1.36***	0.20	0.26	-0.53**	0.17	N/A
Asian	-1.49**	0.46	0.23	-0.67*	0.27	N/A
Other Race/Ethnic Group	-1.31***	0.29	0.27	-0.52***	0.15	N/A
Kindergarten SES	0.37**	0.13	1.45	0.30*	0.12	N/A
Age at Fall Kindergarten Assessment	0.09***	0.02	1.09	0.04*	0.02	N/A
Constant	-5.58***	1.13	0.004	-3.05*	1.04	N/A

Note. [A]Unweighted model uses logit model. [B]Weighted model uses probit model. e^{β} = exponetiated **β** (log odds). * $p < .05$; ** $p < .01$; *** $p < .001$.

Growth Trajectory of the Full Sample

Before conducting the multiple group LGM, two latent growth curves with different time scores were fit to determine the most appropriate pattern of growth in reading achievement in the full sample. The first was a linear growth model. In this model, children were predicted to change in a linear fashion with equal growth each year. Both unweighted and weighted results indicated misfit (unweighted: χ^2 = 1322.35, df = 16, RMSEA = 0.28, CFI = 0.09, TLI = 0.15; weighted: RMSEA = 0.29). Please note that a chi-square test of model fit along with the CFI and TLI are not available in MPlus when using replicate weights. Based on these results, this linear growth model did not represent the growth in reading achievement over time, such that students did not grow at the same rate from year to year.

Given the misfit, modification indices were examined. The modification indices display the amount of decrease in χ^2 that is expected if certain fixed parameters were freely estimated (Muthén & Muthén, 2012). These indices suggested that freeing the residual covariances/regressions of the reading IRT scale scores would improve model fit. Therefore, the residual covariances/regressions were set to be freely estimated using a directional regression path (i.e., spring kindergarten on fall kindergarten, first grade on spring kindergarten, third grade on first grade, fifth grade on third grade, eighth grade on fifth grade). This new unweighted linear model, χ^2 = 69.62, df = 11, RMSEA = 0.07, CFI = 0.96, TLI = 0.94, was a significant improvement over the linear model without the residual covariances/regressions ($\Delta\chi^2$ = 1330.29, Δdf = 5, $p < .001$). The RMSEA for this new weighted linear model was 0.09, which also was an improvement over the linear model without the residual covariances/regression; however, the Satorra-Bentler scaled chi-square difference test could not be computed, because the chi-square test of model fit is not available in MPlus when using replicate weights. Despite the improvement in these linear models, model fit was not strong enough, especially given that RMSEA > 0.06, and its p-value was significant (i.e., $p = 0.02$). Therefore, a linear model even with the residual covariances/regressions was not appropriate for the given data.

Accordingly, due to misfit of the linear model, an unconditional latent basis growth model was estimated. In this non-linear model, the first shape coefficient was fixed to zero (i.e., fall kindergarten) and the last to one (i.e., eighth grade spring), and the four middle shape coefficients were estimated freely. Results of the unweighted and weighted models indicated model misfit (unweighted: $\chi^2 = 143.73$, df = 12, RMSEA = 0.10, CFI = 0.91, TLI = 0.89; weighted: RMSEA = 0.12). Still, the unweighted latent basis model fit substantially better than the linear model without residual covariances/regressions ($\Delta\chi^2 = 1220.74$, Δdf = 4, $p < .001$).

Because fit was improved in the latent basis growth model, modification indices were examined. As in the linear model, residual covariances/regressions were added into the model using a directional regression path (i.e., spring kindergarten on fall kindergarten, first grade on spring kindergarten, third grade on first grade, fifth grade on third grade, eighth grade on fifth grade). This model fit the data well, $\chi^2 = 10.52$, df = 7, RMSEA = 0.02, CFI = 0.998, TLI = 0.995, and it was a significant improvement over the latent basis model without the residual covariances/regressions ($\Delta\chi^2 = 124.89$, Δdf = 5, $p < .001$).

As a result of the strong fit of the latent basis model with the residual covariances/regressions, this model was accepted as the final unconditional model that best explained growth in reading achievement in the current sample of children with RD and LRA. The four estimated basis coefficients equaled 0.03, 0.08, 0.59, and 0.87. The first estimated basis coefficient of 0.03 was not significant ($p > .05$), indicating that the growth in reading achievement from the fall of kindergarten to the spring of kindergarten was not statistically meaningful. The three other basis coefficients were significant (i.e., first grade $p < .05$, third grade $p < .01$, fifth grade $p < .001$). These basis coefficients suggest that 8% of the growth occurred between the fall of kindergarten and the spring of first grade, 59% of the growth occurred between the fall of kindergarten and the spring of third grade, and 87% of the growth occurred between the fall of kindergarten and the spring of fifth grade. Interestingly, while 59% of growth occurred between the fall of kindergarten and the spring of third grade, 51% of this group occurred between the spring of first grade and the spring of third grade. In other words, over half of the growth in reading achievement occurred between the spring of first grade and the spring of third grade, with another 28% of growth occurring between the spring of third grade and the spring of fifth grade, and the remaining 13% occurring between the spring of fifth grade and the spring of eighth grade. Please see Figure 4 for the predicted trajectories of the full sample.

In the unconditional latent basis model with residual covariances/regressions, the covariance and correlation between the initial level and total change was not significant, which indicates that initial level and total growth were not significantly related (e.g., children who started with lower reading achievement did not change more than those who started with higher reading achievement; children with lower reading achievement did not change less than those with higher reading achievement; $p = .84$). That is, neither a “catch-up effect,” in which the lowest performing children would catch up to their peers, nor a “fan-spread effect,” in which rank order stability would be maintained over time, occurred here. There was, however, significant variation in the initial starting level ($p < .001$), but not in the total growth ($p = .18$), which suggests that there were significant individual differences in the reading development in the fall of kindergarten, but that there were not significant individual differences in terms of total growth over time. In other words, there was a lot of variability where children started in the growth process, but there was not a lot of variability in their rate of change over time. The parameter estimates and fit statistics for the unweighted models appear in Table 31.

Although there was not significant variation in the total growth from kindergarten to eighth grade, there appeared to be great variation in reading achievement scores at the end of the eighth grade. Figure 5 displays the spaghetti plots with varying numbers of participants. From visual analysis of these plots, one can see that there was less variation in the reading achievement scores at the start of kindergarten, although still significant, and by the end of the eighth grade, this variation, or fan-spread effect, was highly evident. It is possible that these large variances in reading achievement scores over time contributed to the non-positive definite covariance matrices (i.e., negative slope and intercept variances) in some of both the unweighted and weighted analyses discussed below.

This unconditional latent basis model with residual covariances/regressions was also tested on the weighted data (see Table 32). The RMSEA of the weighted model was 0.06, which is higher than the RMSEA of the unweighted model, RMSEA = 0.02. Also of note, this model produced a negative slope variance (-3.37, p = .33), which suggests that this model was not the best for the weighted data. The negative slope variance also influenced the estimates for basis coefficients, such that one of the estimates was negative and two of the estimates were greater than two, which are neither plausible nor interpretable estimates. Accordingly, it was inappropriate to use the weighted model to draw conclusions about the growth trajectory in the full sample. Although conclusions about the longitudinal trajectory cannot be drawn from the weighted unconditional latent basis model with residual covariances/regressions, the results from the unweighted model may be somewhat applicable to the national population, given the limited differences between the weighted and unweighted samples, as described above.

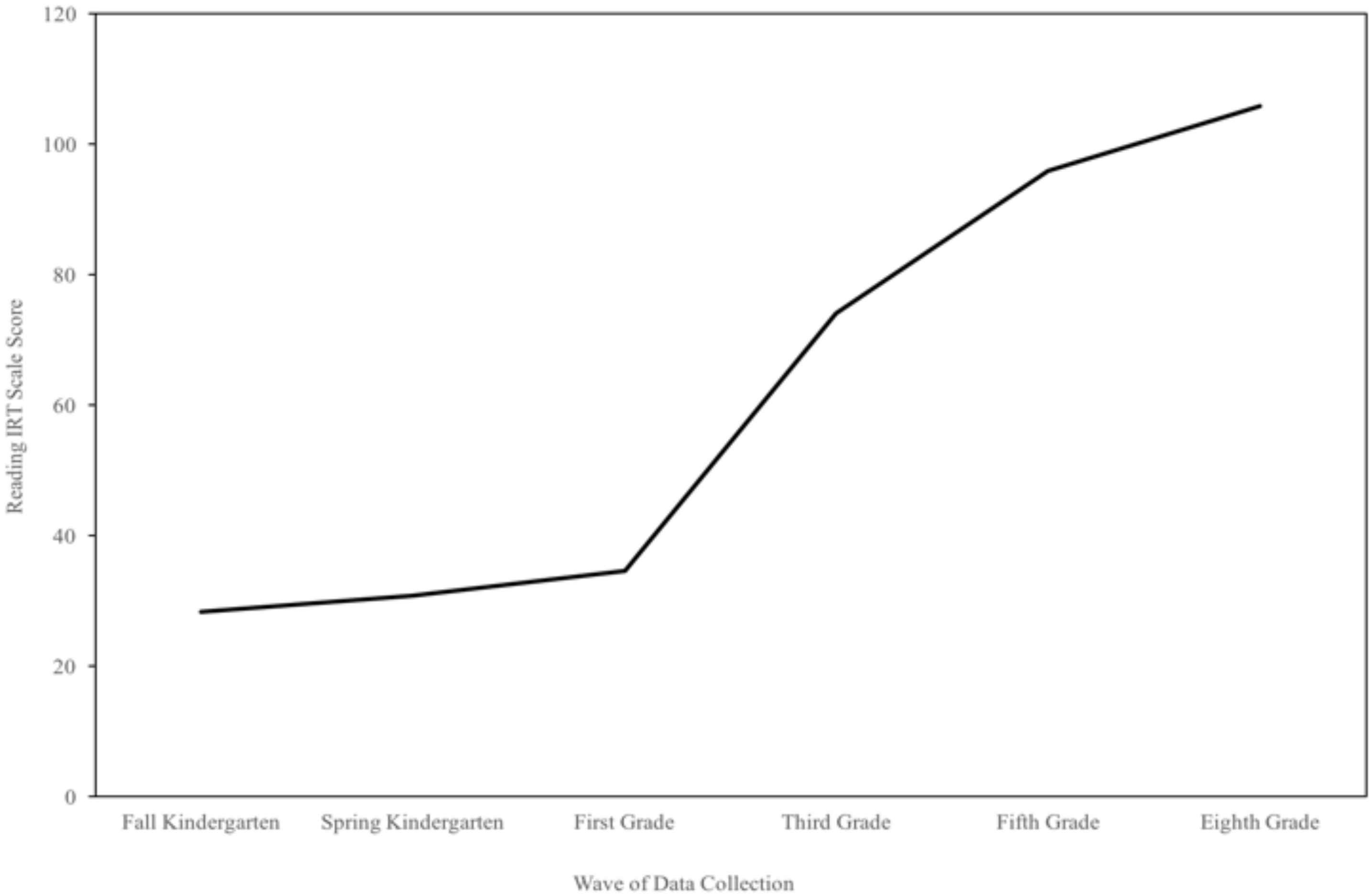

Figure 4. Predicted Reading Growth Trajectory of Full Sample

Table 31

Parameter Estimates and Fit Statistics for Unweighted Unconditional LGMs

Parameter	Linear Growth	Linear Growth with Residual Covariances/ Regressions	Latent Growth	Latent Growth with Residual Covariances/Regressions
Fixed Effects				
Mean Initial Level	28.45 (.14)***	28.33 (.14)***	28.34 (.14)***	28.32 (.14)***
Mean Slope	15.76 (.10)***	10.26 (.73)***	109.46 (.80)***	77.40 (25.66) **
Basis Coefficients	=0,	=0,	=0,	=0
	=.5,	=.5,	0.07 (.00)***,	0.03 (.02)
	=1.5,	=1.5,	0.22 (.00)***,	0.08 (.02)*
	=3.5,	=3.5,	0.56 (.01)***,	0.59 (.20)**
	=5.5,	=5.5,	0.81 (.01)***,	0.87 (.25)***
	=8.5	=8.5	=1	=1
Random Effects				
Initial Level Variance	13.37 (.92)***	12.27 (.89)***	13.52 (.91)***	10.26 (1.10)***
Slope Variance	5.42 (.48)***	2.98 (.43)***	294.40 (20.01)***	158.64 (117.33)
Initial Level/Slope Correlation	0.04 (.051)	-0.11 (.58)	0.00 (.05)	0.02 (.08)
Initial Level/Slope Covariance	0.30 (.43)	-0.64 (.35)	0.23 (2.93)	0.64 (3.26)
Fit Statistics				
χ^2 / df	1322.34 / 16	69.62 / 11	143.73 / 12	10.52 / 7
RMSEA / CI[A]	0.28 / .26-.29	0.07 / .06-.09	0.10 / .09-.12	0.02 / .00-.05
CFI	0.09	0.96	0.91	0.998
TLI	0.15	0.94	0.89	0.995

Note. [A]CI = confidence interval. Values preceded by = indicate that the parameter was fixed to the specified value. Values in parentheses are standard errors. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index. * $p < .05$; ** $p < .01$; *** $p < .001$.

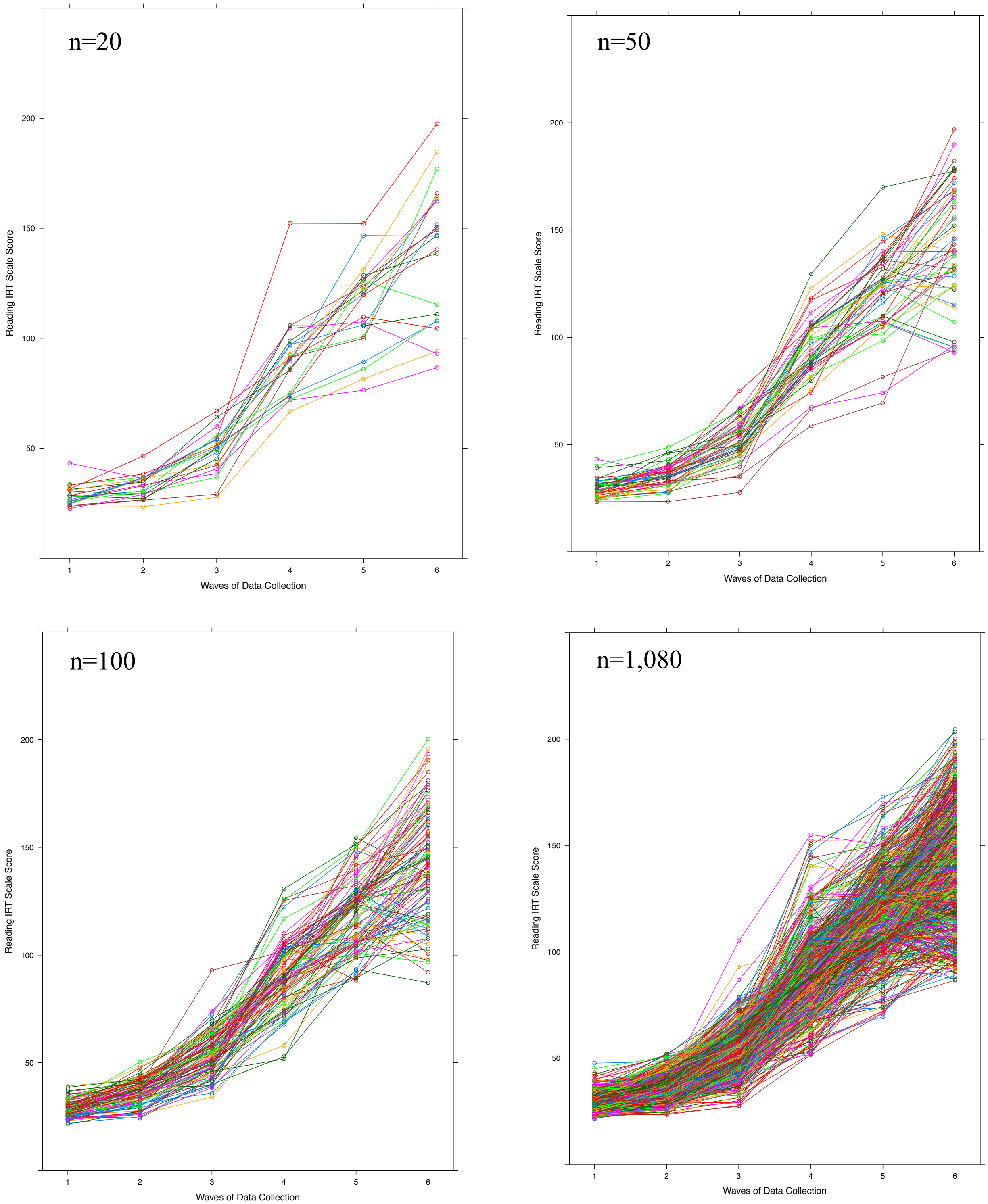

Figure 5. Spaghetti Plots of Full Sample

Table 32

Parameter Estimates and Fit Statistics for Weighted Unconditional LGMs

Parameter	Linear Growth	Linear Growth with Residual Covariances/ Regressions	Latent Growth	Latent Growth with Residual Covariances/Regressions
Fixed Effects				
Mean Initial Level	28.26 (.23)***	11.21 (1.23)***	28.18 (.22)***	28.17 (0.23)***
Mean Slope	15.70 (.18)***	2.61 (.79)***	107.08 (1.43)***	12.27 (11.14)
Basis Coefficients	=0,	=0,	=0	=0
	=.5,	=.5,	0.07 (.00)***	-0.09 (.18)
	=1.5,	=1.5,	0.22 (.01)***	0.27 (0.35)
	=3.5,	=3.5,	0.57 (.01)***	2.88 (2.33)
	=5.5,	=5.5,	0.82 (.01)***	2.02 (1.13)
	=8.5	=8.5	=1	=1
Random Effects				
Initial Level Variance	12.54 (1.29)***	11.21 (1.23)***	12.63 (1.30)***	8.01 (1.64)***
Slope Variance	5.48 (.68)***	2.61 (.79)**	285.94 (30.17)***	-3.37 (3.48)
Initial Level/Slope Correlation	0.08[B]	-0.03[B]	0.05[B]	N/A
Initial Level/Slope Covariance	0.69 (.55)	-0.17 (.43)	3.15 (3.54)	-0.74 (1.58)
Fit Statistics				
χ^2 / df	N/A	N/A	N/A	N/A
RMSEA / CI[A]	0.29 / .28-.31	0.09 / .07-.10	0.12 / .10-.13	0.06 / .05-.09
CFI	N/A	N/A	N/A	N/A
TLI	N/A	N/A	N/A	N/A

Note. [A]CI = confidence interval. [B]Standard error and significance value not available. Values preceded by = indicate that the parameter was fixed to the specified value. Values in parentheses are standard errors. N/A = not available; χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index. * $p < .05$; ** $p < .01$; *** $p < .001$.

Covariates and Growth Trajectory in the Full Sample

With the longitudinal nature of reading achievement established in the unconditional model, a conditional latent basis growth model with residual covariances/regressions was fit. The time invariant covariates included in this model were gender, race, SES at fall kindergarten assessment, and age at fall kindergarten assessment. Dummy variables were created for the categorical variables (i.e., gender, race). In this conditional model, the purpose was to examine the nature of the relationship among the covariates, the growth factors, and reading achievement outcome variables across the entire sample.

The conditional latent basis model with residual covariances/regressions fit the data well, $\chi^2 = 61.41$, df = 35, RMSEA = 0.03, CFI = 0.99, TLI = 0.98. The basis coefficients of the conditional model were similar to those of the unconditional model that was described above. Results regarding the random effects were also similar to those of the unconditional model, such that there was significant variation in the initial starting level ($p < .001$), but not in the total change ($p = .10$). The covariance and correlation between the initial level and total change were also not significant, which indicated that children who started with lower reading achievement did not change more than those who started with higher reading achievement and vice versa (covariance: $p = .44$; correlation: $p = .81$).

Perhaps the most important results of the unweighted conditional model were the relationships between the covariates and the initial level and the total change over time. First, in terms of the relationship between the covariates and the initial level, all of the covariates significantly predicted the initial level of reading achievement, except for the race/ethnicity category of African American children. In other words, Hispanics, Asians, and children from other diverse race/ethnicities all started with significantly lower reading achievement in the fall of kindergarten in comparison to White children. African American children, however, did not display reading achievement that was significantly different from White children in the fall of kindergarten (p = .39). Additionally, girls demonstrated significantly higher reading achievement in the fall of kindergarten than did boys. Regarding SES, for every one unit increase in SES, students' reading achievement in the fall of kindergarten increased by 1.04. That is, children with higher SES in the fall of kindergarten had significantly higher reading achievement scores in the fall of kindergarten. Finally, for every one unit increase in age, children scored 0.24 units higher on the direct reading achievement test in the fall of kindergarten, such that older children performed better on the reading achievement test in the fall of kindergarten than did younger children.

Second, four out of the seven covariates significantly predicted the total change, or growth, over time. To start, gender did not significantly predict the total change over time, such that boys and girls made similar gains from kindergarten through eighth grade (p = .12). While girls had initially higher reading achievement in the fall of kindergarten, boys and girls made similar progress in reading achievement over time. Interestingly, Hispanic children and Asian children did not differ from White children in terms of their total growth over time (Hispanic: p = .10; Asian: p = .87), despite starting at significantly lower starting points than White children in the fall of kindergarten. For African American children, however, despite having similar reading achievement scores in the fall of kindergarten in comparison to White children, they made significantly less gains in their reading achievement than did White children over time. For children from other diverse races/ethnicities, they started with significantly lower reading achievement scores than White children and made significantly less gains over time in comparison to White children. As for SES, for every one unit increase in SES, the total change increased by 4.45, which suggests that children with higher SES not only started with significantly higher reading achievement in the fall of kindergarten, but they also continued to make more progress than children with lower SES. Regarding age, for every one unit increase in age, the total change decreased by 0.49, indicating that although older children started with significantly higher reading achievement in the fall of kindergarten, they made less progress over time in comparison to younger children.

In order to examine the relationship between the covariates and students' reading achievement scores in eighth grade, the latent basis model with residual covariances/regressions was run using eighth grade as the initial level, or intercept. The kindergarten fall time point was fixed to -1; the kindergarten spring, first grade, third grade, and fifth grade time points were freely estimated; and the eighth-grade time point was fixed to zero. The zero point of the time scale determines the time at which to interpret the parameter estimates (Rogosa & Willett, 1985), so by fixing the eighth-grade time point to zero, it can be interpreted as the intercept.

With eighth grade as the intercept, results indicated that girls demonstrated significantly higher reading achievement than boys in eighth grade. For race/ethnicity, African Americans, Hispanics, and children from other diverse races/ethnicities had significantly lower reading achievement than White children in eighth grade, while there were no significant differences between the reading achievement scores of Asian and White children at the end of eighth grade ($p = .65$). For SES, children with higher levels of SES had significantly higher reading achievement scores at the end of eighth grade, and lastly, age did not significantly predict reading achievement scores at the end of eighth grade ($p = .06$), which suggests that there were no differences in the performance of younger and older children by the end of eighth grade.

Fit statistics and parameter estimates of the conditional model with the weighted data were also evaluated for the models with both the initial level set at the fall of kindergarten and the initial level set at eighth grade. For these weighted models, the RMSEA was equal to 0.05; however, the basis coefficients were not similar, or comparable to those of the weighted and unweighted unconditional models, such that the final two basis coefficients were greater than one. Basis coefficients greater than one indicated that this latent basis model did not capture the shape of growth over time adequately. Consequently, the relationships between the covariates and the initial level and shape of growth over time in the weighted model were not interpreted. As such, the interpretation of the relationships between the covariates and the initial level and shape of growth over time were only made about the unweighted conditional model, and although this limits the generalizability of the results, it was the most appropriate decision given the inadequate results of the weighted conditional model.

In sum, only the results of the unweighted conditional model were interpreted in detail. Accordingly, there were interesting patterns of relationships between the covariates and children's reading achievement in the fall of kindergarten, how much progress they made over time, and where they ended up in eighth grade. Girls had significantly higher reading achievement compared to boys in the fall of kindergarten and at the spring of eighth grade, but boys and girls made similar progress in their reading achievement between kindergarten and eighth grade. African American children had similar reading achievement compared to White children in the fall of kindergarten, but they progressed significantly less and had significantly lower scores in eighth grade in comparison to White children. Hispanic children started and ended significantly lower than White children, but they made gains that were similar to White children. Asian children also started lower than White children and made gains that were similar to White children, but they were the only race/ethnicity group whose reading achievement did not differ from White children at the end of eighth grade. Children from other diverse races/ethnicities started and ended significantly lower than White children, and they also made significantly less gains compared to White children. For SES, children with higher SES started and ended significantly higher than children with lower SES, and they made significantly greater gains over time than did children with lower SES. Lastly for age, older children had significantly higher reading achievement in the fall of kindergarten, but older children made significantly less gains over time as compared to younger children; by the spring of eighth grade, there were no significant differences in the reading achievement scores based on age. Please see Table 33 for the parameter estimates and fit statistics for the unconditional conditional models.

Table 33

Parameter Estimates and Fit Statistics for Unweighted Conditional LGMs

Parameter	Latent Growth with Residual Covariances/Regressions (Kindergarten Intercept)	Latent Growth with Residual Covariances/Regressions (Eighth Grade Intercept)
Fixed Effects		
Mean (Intercept) Initial Level	12.78 (1.96)***	137.83 (30.23)***
Mean (Intercept) Slope	125.04 (30.19)***	125.04 (30.18)***
Basis Coefficients	=0	=-1
	0.03 (.02)	-0.97 (.02)***
	0.08 (.03)*	-0.93 (.03)***
	0.52 (.14)***	-0.48 (.14)**
	0.75 (.16)***	-0.25 (.16)
	=1	=0
Initial Level on Covariates		
Female	0.74 (.24)**	2.74 (1.23)*
African American	-0.28 (.32)	-11.85 (3.20)***
Hispanic	-1.26 (.33)***	-4.30 (1.78)*
Asian	-1.72 (.74)*	-1.25 (2.74)
Other Race/Ethnic Group	-1.68 (.40)***	-11.60 (3.13)***
Kindergarten SES	1.04 (.18)***	5.49 (1.57)***
Age at Fall Kindergarten Assessment	0.24 (.03)***	-0.26 (.14)
Shape on Covariates		
Female	2.00 (1.27)	2.00 (1.27)
African American	-11.57 (3.21)***	-11.57 (3.21)***
Hispanic	-3.04 (1.86)	-3.04 (1.86)
Asian	0.47 (2.80)	0.47 (2.80)
Other Race/Ethnic Group	-9.91 (3.13)**	-9.91 (3.13)**
Kindergarten SES	4.45 (1.58)**	4.45 (1.58)**
Age at Fall Kindergarten Assessment	-0.49 (.14)***	-0.49 (.14)***

Table 33 (cont'd)

Random Effects		
Initial Level Residual Variance	8.11 (.97)***	137.83 (30.23)***
Slope Residual Variance	161.50 (99.38)	125.04 (30.18)***
Initial Level/Slope Correlation	-0.02 (.08)	0.97 (.02)***
Initial Level/Slope Residual Covariance	-2.23 (2.88)	159.27 (99.88)
Fit Statistics		
χ^2 / df	61.41 / 35	61.41 / 35
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .02-.04
CFI	0.99	0.99
TLI	0.98	0.98

Note. [A]CI = confidence interval. Values preceded by = indicate that the parameter was fixed to the specified value. Values in parentheses are standard errors. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index. * $p < .05$; ** $p < .01$; *** $p < .001$.

Model Fit in RD and LRA Groups

Before conducting the multiple group LGM, it was necessary to determine that the unconditional latent basis model fit each group adequately. Results of the unweighted unconditional model with residual covariances/regressions for the RD group indicated adequate fit, χ^2 = 8.18, df = 7, RMSEA = 0.02, CFI = 0.998, TLI = 0.996. Similar results were found for the LRA group, χ^2 = 8.74, df = 7, RMSEA = 0.02, CFI = 0.998, TLI = 0.995. Fit of the weighted models for the RD and LRA groups were poorer than the unweighted models, but perhaps still mediocre, given that the lower value of the RMSEA confidence interval for the RD group was less than 0.05 and for the LRA group was equal to 0.05 (RD: RMSEA = 0.06, confidence interval [CI] = 0.10-.010; LRA: RMSEA = 0.07, CI = 0.05-0.09). Also of note, the weighted unconditional model for the RD group produced a negative slope variance (-10.07, p = .55), suggesting that this model did not fit the weighted data appropriately.

Fit of the conditional models for each group were also considered. Results of the unweighted conditional latent basis model with residual covariances/regressions for the RD group indicated adequate fit, χ^2 = 38.32, df = 35, RMSEA = 0.02, CFI = 0.996, TLI = 0.99. Results of the unweighted conditional latent basis model with residual covariances/regressions for the LRA group also indicated adequate fit, but the fit was slightly worse in comparison to the RD group, χ^2 = 62.53, df = 35, RMSEA = 0.03, CFI = 0.97, TLI = 0.95. These results suggest that the unweighted unconditional and conditional latent basis models with residual covariances/regressions fit each group well, and, thus, a multiple group LGM would be appropriate. Please see Table 34 for fit statistics of the unweighted unconditional and conditional models for each group.

In the weighted conditional models, fit for both the RD and LRA groups was slightly worse than in the unweighted models. The RMSEA for the RD group was 0.07 (CI = 0.05-0.09), and the RMSEA for the LRA group was 0.05 (CI = 0.04-0.06). The weighted conditional model for the LRA group produced a negative slope variance (-0.47, p = .97). Given the negative slope variance in the LRA group, the parameter estimates for the weighted conditional models were not interpreted; rather, conclusions about group differences in the conditional model were interpreted using the unweighted models.

Accordingly, in addition to examining the fit statistics of each model in the unweighted unconditional and conditional models, the parameter estimates of the conditional model for each group separately were analyzed (see Table 35). Parameter estimates for the conditional model were examined because the conditional model was used to conduct the multiple group analyses. Interestingly, it appeared that the RD group had a slightly lower initial starting value in comparison to the LRA group, and the RD group also appeared to have a greater mean total change compared to the LRA group.

Regarding the basis coefficients, or the pattern of growth over time in the unweighted model, the RD and LRA groups appeared to experience similar patterns of growth until the first-grade time point. In other words, for the RD group, 12% of the growth in reading achievement occurred between the fall of kindergarten and the spring of first grade, and for the LRA group, 10% of the growth in reading achievement occurred between the fall of kindergarten and the spring of first grade. The growth from the fall of kindergarten through the spring of first grade, however, was not statistically significant in either group, but after first grade, all remaining growth was statistically significant. Then, from the spring of first grade until the spring of eighth grade, the RD group experienced 88% of the growth in reading achievement. More specifically, the RD group experienced 27% of growth between the spring of first grade and the spring of third grade, 19% of growth between the spring of third grade and the spring of fifth grade, and 42% of growth between the spring of fifth grade and the spring of eighth grade. For the LRA group, students experienced growth greater than 100% between the spring of first grade and the spring of fifth grade, before going back down at the final eighth-grade time point. In other words, the LRA group experienced increased acceleration in reading growth at the end of first grade through the end of fifth grade, but then experienced a deceleration in reading growth from the end of fifth grade to the end of eighth grade. However, the basis coefficients for the LRA group (i.e., estimates greater than one) suggest that this model may not be the best fit for the data. Thus, it is important that the basis coefficients for the LRA group are not overly interpreted, because it is difficult to say exactly what these participants' pattern of growth looked like, especially after the first-grade time point. Despite being unsure of the LRA group's exact shape of reading development over time, the two groups appeared to have similar initial starting values and patterns of change, and, therefore, the multiple group analyses must be conducted in order to determine if there were statistically significant differences.

Table 34

Fit Statistics for Unweighted Unconditional and Conditional LGMs by Group

Fit Statistics	**RD Group Unconditional Model**	**LRA Group Unconditional Model**	**RD Group Conditional Model**	**LRA Group Conditional Model**
χ^2 / df	8.18 / 7	8.74 / 7	38.32 / 35	62.53 / 35
RMSEA / CI^A	0.02 / .00-.08	0.02 / .00-.05	0.02 / .00-.05	0.03 / .02-.04
CFI	0.998	0.998	0.996	0.97
TLI	0.996	0.995	0.994	0.95

Note. ^{A}CI = confidence interval. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index. **p* < .05; ** *p* < .01; *** *p* < .001.

Table 35

Parameter Estimates for Unweighted Conditional Latent Basis Models for Each Group

Parameter	**RD Group**	**LRA Group**
Fixed Effects		
Mean (Intercept) Initial Level	7.53 (3.60)*	13.68 (2.40)***
Mean (Intercept) Slope	103.46 (41.01)*	81.82 (21.00)***
Basis Coefficients	=0	=0
	0.05 (.03)	0.04 (.04)
	0.12 (.06)	0.10 (.06)
	0.39 (.12)**	1.05 (.28)***
	0.58 (.17)***	1.33 (.27)***
	=1	=1
Initial Level on Covariates		
Female	0.62 (.46)	0.74 (.28)**
African American	-0.43 (.80)	-0.62 (.38)
Hispanic	-2.25 (.55)***	-1.36 (.40)**
Asian	-6.09 (1.78)**	-1.62 (.78)*
Other Race/Ethnic Group	-1.00 (.74)	-1.98 (.48)***
Kindergarten SES	1.81 (.36)***	0.84 (.21)***
Age at Fall Kindergarten Assessment	0.31 (.05)***	0.23 (.04)***
Shape on Covariates		
Female	2.28 (2.93)	0.77 (.73)
African American	-13.83 (6.18)*	-4.92 (1.64)**
Hispanic	-6.41 (4.58)	0.08 (.90)
Asian	2.29 (7.20)	2.02 (1.71)
Other Race/Ethnic Group	-19.46 (7.91)*	-3.27 (1.45)*
Kindergarten SES	9.06 (4.01)*	1.47 (.74)
Age at Fall Kindergarten Assessment	-0.20 (.27)	-0.41 (.12)**
Random Effects		
Initial Level Residual Variance	10.61 (1.80)***	6.75 (1.23)***
Slope Residual Variance	260.26 (231.45)	27.20 (22.24)

Table 35 (cont'd)

Initial Level/Slope Correlation	0.01 (.13)	-0.10 (.13)
Initial Level/Slope Residual Covariance	-9.95 (6.38)	-0.74 (1.85)

Note. Values preceded by = indicate that the parameter was fixed to the specified value. Values in parentheses are standard errors. * p < .05; ** p <.01; *** p <. 001.

Note About Weighted Models

The results of the final three research questions related to group differences in initial levels of reading achievement, shape of development over time, and eighth-grade reading outcomes are provided in the subsequent sections. These questions, however, were only answered using the unweighted data. Although using only the unweighted data limits the generalizability of the results, the weighted data were not used for these analyses for two reasons: (1) in MPlus, chi-square statistics as well as Wald tests are not available when using replicate weights, which then does not allow for the comparison of group differences, and (2) a negative slope variance existed for the LRA group in the weighted conditional model, and although this did not automatically invalidate the data, it suggested that the model was not the best fit for the data. Because a negative slope variance did not occur in either group when using the unweighted data in the preliminary group analyses described above, the unweighted data were used to conduct the multiple group LGMs.

Research Question 2: Group Comparison of Initial Level of Reading Achievement

To examine whether there were group differences in the initial starting level of reading achievement at the fall of kindergarten, a multiple group analysis was conducted utilizing the unweighted conditional latent basis models with residual covariances/regressions. First, an unconstrained model was run, in which the initial levels were not constrained to be equal. The parameter estimates for each group in the unconstrained model can be found in Table 35, as the parameter estimates for the unconstrained model are the same as when the conditional model was run for each group separately. Second, a constrained model was run, in which the initial levels were constrained to be equal. Please see Table 36 for parameter estimates of the constrained model.

With the unconstrained and constrained models established, the Satorra-Bentler scaled chi-square difference test was used to determine if significant differences between the two groups' initial starting levels existed. Please see Table 37 for fit statistics of the unconstrained and constrained models. Results indicated no significant group differences in the initial starting levels ($\Delta\chi^2 = 2.10$, $\Delta df = 1$) at the .05 level. That is, the difference between the chi-square values (2.10) was not greater than the chi-square statistic for one degree of freedom at the .05 level (3.84). Non-significant results of the Wald test also supported the finding of a non-significant difference in the initial level of reading achievement between the two groups ($t = 1.72$, df = 1, $p = .19$). Therefore, these results suggest that the RD group and the LRA group had similar levels of reading achievement in the fall of kindergarten (see Figure 6).

Table 36

Parameter Estimates for Initial Level Constrained Model

Parameter	**RD Group**	**LRA Group**
Fixed Effects		
Mean (Intercept) Initial Level	11.58 (2.02)***	11.58 (2.02)***
Mean (Intercept) Slope	99.80 (42.60)*	82.18 (21.01)***
Basis Coefficients	=0	=0
	0.05 (.03)	0.05 (.04)
	0.12 (.06)	0.11 (.06)
	0.38 (.12)**	1.07 (.28)***
	0.57 (.17)**	1.33 (.26)***
	=1	=1
Initial Level on Covariates		
Female	0.52 (.46)	0.76 (.29)**
African American	-0.37 (.80)	-0.54 (.38)
Hispanic	-2.29 (.55)***	-1.34 (.40)**
Asian	-6.12 (1.81)**	-1.60 (.78)*
Other Race/Ethnic Group	-1.11 (.76)	-2.00 (.49)***
Kindergarten SES	1.79 (.36)***	0.86 (.21)***
Age at Fall Kindergarten Assessment	0.26 (.03)***	0.26 (.03)***
Shape on Covariates		
Female	2.45 (2.97)	0.77 (.73)
African American	-14.07 (6.27)*	-4.91 (1.63)**
Hispanic	-6.50 (4.68)	0.09 (.89)
Asian	2.25 (0.76)	2.03 (1.69)
Other Race/Ethnic Group	-19.57 (8.04)*	-1.39 (1.44)*
Kindergarten SES	9.22 (4.10)*	1.44 (.73)
Age at Fall Kindergarten Assessment	-0.13 (.27)	-0.43 (.12)***
Random Effects		
Initial Level Residual Variance	10.65 (1.80)***	7.03 (1.18)***
Slope Residual Variance	269.52 (241.90)	26.01 (21.85)

Table 36 (cont'd)

Initial Level/Slope Correlation	0.02 (.13)	-0.14 (.12)
Initial Level/Slope Residual Covariance	-10.10 (6.49)	-1.01 (1.79)

Note. Values preceded by = indicate that the parameter was fixed to the specified value. Values in parentheses are standard errors. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 37

Fit Statistics for Multiple Group LGM for Initial Level

Fit Statistics	**Unconstrained Model**	**Constrained Model**
χ^2 / df	101.87 / 70	104.12 / 71
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .02-.04
CFI	0.98	0.98
TLI	0.97	0.97

Note. [A]CI = confidence interval. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index. * $p <.05$; ** $p <.01$; *** $p < .001$.

Research Question 3: Group Comparison of Shape of Reading Achievement

Given that there were no significant differences between the two groups' reading achievement at the start of kindergarten, it was important to examine whether the overall pattern, or shape of change from kindergarten to eighth grade differed between the two groups using the unweighted data. First, an unconstrained model was run, in which the basis coefficients were not constrained to be equal. This was the same unconstrained model that was used to answer Research Question 2 (see Table 35 for parameter estimates). Second, a constrained model was run, in which the basis coefficients were constrained to be equal. Please see Table 38 for parameter estimates of the constrained model.

With the unconstrained and constrained models established, the Satorra-Bentler scaled chi-square difference test was used to determine if significant differences between the two groups' shapes existed. Please see Table 39 for fit statistics of the unconstrained and constrained models. Results indicated significant group differences in the shapes ($\Delta\chi^2 = 13.46$, $\Delta\text{df} = 4$, $p < .01$). That is, the difference between the chi-square values (13.46) was greater than the chi-square statistic for four degrees of freedom at the .05 level (9.49). A Wald test could not be computed, because MPlus does not allow for equality constraints to be placed in a growth statement. Therefore, these results suggest that the RD group and the LRA group had different shapes, or patterns in their reading achievement growth from kindergarten through eighth grade.

Related to the group differences in the basis coefficients, or pattern of reading development over time, the RD and LRA groups appeared to experience similar patterns of growth until the first-grade time point, such that around 10% of the growth in reading achievement occurred between the fall of kindergarten and the spring of first grade. Also for both groups, growth was not statistically significant until after the first-grade time point. Then, from the spring of first grade until the spring of eighth grade, the RD group experienced 88% of the growth in reading achievement. More specifically, the RD group experience 27% of growth between the spring of first grade and the spring of third grade, 19% of growth between the spring of third grade and the spring of fifth grade, and 42% of growth between the spring of fifth grade and the spring of eighth grade.

For the LRA group, students experienced growth greater than 100% between the spring of first grade and the spring of fifth grade, before going back down at the final eighth-grade time point. In other words, the LRA group experienced increased acceleration in reading growth at the end of first grade through the end of fifth grade, but then experienced a deceleration, or reduction, in reading growth from the end of fifth grade to the end of eighth grade. Of note, between the spring of third grade and the spring of fifth grade, it appeared that the LRA group still experienced growth, but it was slightly slower than the growth from the spring of first grade to the spring of third grade. Further, the basis coefficients for the LRA group (i.e., estimates greater than one) suggest that this model was not the best fit for the data. Although the LRA group's basis coefficients and subsequent shape of reading development over time were less interpretable, it is still arguable that the RD and LRA groups had significantly different patterns, or shapes of reading achievement growth over time, but exactly what these differences were remains unclear (see Figure 6).

Table 38

Parameter Estimates for Basis Coefficients Constrained Model

Parameter	RD Group	LRA Group
Fixed Effects		
Mean (Intercept) Initial Level	8.17 (3.37)*	13.43 (2.45)***
Mean (Intercept) Slope	42.98 (16.31)**	91.48 (23.38)***
Basis Coefficients	=0	=0
	0.04 (.04)	0.04 (.04)
	0.11 (.05)*	0.11 (.05)*
	0.93 (.25)***	0.93 (.25)***
	1.22 (.25)***	1.22 (.25)***
	=1	=1
Initial Level on Covariates		
Female	0.69 (.44)	0.73 (.29)*
African American	-0.57 (.80)	-0.60 (.38)
Hispanic	-2.22 (.54)***	-1.37 (.40)**
Asian	-5.40 (1.65)**	1.65 (.78)*
Other Race/Ethnic Group	-1.09 (.71)	-1.99 (.49)***
Kindergarten SES	1.89 (.34)***	0.84 (.21)***
Age at Fall Kindergarten Assessment	0.30 (.05)***	0.23 (.04)***
Shape on Covariates		
Female	0.18 (1.29)	0.90 (.84)
African American	-5.66 (2.72)*	-5.56 (1.89)**
Hispanic	-1.92 (1.79)	0.05 (.99)
Asian	2.39 (3.44)	2.19 (1.90)
Other Race/Ethnic Group	-8.44 (3.40)*	-3.67 (1.54)*
Kindergarten SES	2.86 (1.78)	1.66 (.82)*
Age at Fall Kindergarten Assessment	-0.23 (.13)	-0.46 (.14)**
Random Effects		
Initial Level Residual Variance	8.20 (1.59)***	6.98 (1.36)***
Slope Residual Variance	5.81 (31.67)	37.28 (28.26)

Table 38 (cont'd)

Initial Level/Slope Correlation	-0.16 (.36)	-0.11 (.12)
Initial Level/Slope Residual Covariance	-4.84 (3.14)	-1.03 (2.21)

Note. Values preceded by = indicate that the parameter was fixed to the specified value. Values in parentheses are standard errors. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 39

Fit Statistics for Multiple Group LGM for Shape (Basis Coefficients)

Fit Statistics	**Unconstrained Model**	**Constrained Model**
χ^2 / df	101.87 / 70	115.13 / 74
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .02-.04
CFI	0.98	0.98
TLI	0.97	0.97

Note. [A]CI = confidence interval. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index. * p <.05; ** $p < .01$; *** $p < .001$.

Research Question 4: Group Comparison of Eighth Grade Reading Achievement

As described above, the unweighted latent basis model was run with the initial level, or intercept, set to the eighth-grade time point in order to determine whether group differences in reading achievement existed at the end of eighth grade. In this model, the kindergarten fall time point was fixed to -1; the kindergarten spring, first grade, third grade, and fifth grade time points were freely estimated; and the eighth-grade time point was fixed to zero. First, an unconstrained model was run, in which the initial levels (i.e., eighth grade) were not constrained to be equal. This model produced a negative intercept variance (-22.45, p =.16), a negative slope variance (-18.96, p =.08), and a negative covariance (-24.48, p =.07) in the RD group. The negative variances of the growth factors indicated that this model was not the best estimate for the data, and results should be interpreted with caution. Still, it was important to at least explore whether there were group differences in reading achievement at eighth grade. Accordingly, a constrained model, in which the initial levels at eighth grade were constrained to be equal, was run to allow for the continuation of group comparisons. There were no negative variances/covariances in the constrained model. Please see Table 40 and Table 41 for parameter estimates of the constrained model.

With the unconstrained and constrained models established, the Satorra-Bentler scaled chi-square difference test was used to determine if significant differences between the two groups' shapes existed. Due to a negative chi-square difference value, the test could not be completed. Nevertheless, a Wald test was computed, which indicated significant group differences in the reading achievement scores at the end of eighth grade (t = 4.86, df = 1, p < .05). In support of this finding, a Wald test that compared the mean eighth-grade reading achievement scores, in essence an independent samples t-test, revealed similar results (t = 6.09, df=1, $p < .05$). More specifically, the RD group appeared to have significantly outperformed the LRA group at the end of eighth grade (see Figure 6). Therefore, these results suggest that the RD group and the LRA group had different levels of reading achievement in the spring of eighth grade with the RD group scoring significantly higher; however, this model may have produced invalid results due to negative growth factor variances, and results should be interpreted with caution. Please see Table 42 for fit statistics of the unconstrained and constrained models.

Table 40

Parameter Estimates for Unconstrained Model with Eighth Grade as the Initial Level

Parameter	RD Group	LRA Group
Fixed Effects		
Mean (Intercept) Initial Level	32.87 (19.09)	95.50 (21.04)***
Mean (Intercept) Slope	24.00 (19.51)	81.82 (20.99)***
Basis Coefficients	=-1	=-1
	-1.06 (.09)***	-0.96 (.04)***
	-0.58 (.24)*	-0.90 (.06)***
	0.22 (1.03)	0.05 (.28)
	0.45 (.80)	0.33 (.27)
	=0	=0
Initial Level on Covariates		
Female	1.07 (1.08)	1.50 (.71)*
African American	-4.48 (3.01)	-5.53 (1.62)**
Hispanic	-3.05 (1.31)*	-1.28 (.84)
Asian	-2.11 (2.53)	0.40 (1.75)
Other Race/Ethnic Group	-6.94 (4.39)	-5.25 (1.41)***
Kindergarten SES	3.48 (1.86)	2.31 (.74)**
Age at Fall Kindergarten Assessment	0.11 (.12)	-0.18 (.12)
Shape on Covariates		
Female	0.42 (1.20)	0.77 (.73)
African American	-3.85 (3.08)	-4.92 (1.64)**
Hispanic	-0.88 (1.40)	0.08 (.90)
Asian	3.50 (3.28)	2.02 (1.70)
Other Race/Ethnic Group	-5.81 (4.64)	-3.27 (1.45)*
Kindergarten SES	1.58 (1.86)	1.47 (.74)
Age at Fall Kindergarten Assessment	-0.19 (.14)	-0.41 (.12)**
Random Effects		
Initial Level Residual Variance	-22.45 (16.00)	32.47 (22.63)
Slope Residual Variance	-18.96 (10.72)	27.19 (22.23)

Table 40 (cont'd)

Initial Level/Slope Correlation	----------------[A]	0.89 (.08)***
Initial Level/Slope Residual Covariance	-24.48 (13.45)	26.46 (22.49)

Note. [A]Not Available. Values preceded by = indicate that the parameter was fixed to the specified value. Values in parentheses are standard errors. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 41

Parameter Estimates for Initial Level (Eighth Grade) Constrained Model

Parameter	RD Group	LRA Group
Fixed Effects		
Mean (Intercept) Initial Level	99.31 (18.57)***	99.31 (18.57)***
Mean (Intercept) Slope	91.79 (18.93)***	85.59 (18.58)***
Basis Coefficients	=-1	=-1
	-0.94 (.03)***	-0.96 (.04)***
	-0.87 (.05)***	-0.91 (.05)***
	-0.58 (.09)***	0.01 (.23)
	-0.37 (.13)**	0.29 (.23)
	=0	=0
Initial Level on Covariates		
Female	2.70 (2.54)	1.54 (.72)*
African American	-13.17 (5.02)**	-5.77 (1.51)***
Hispanic	-7.82 (3.30)*	-1.31 (.87)
Asian	-3.38 (5.99)	0.44 (1.80)
Other Race/Ethnic Group	-18.81 (5.73)**	-5.42 (1.41)***
Kindergarten SES	10.00 (2.94)**	2.38 (.73)**
Age at Fall Kindergarten Assessment	0.14 (.23)	-0.20 (.11)
Shape on Covariates		
Female	2.08 (2.62)	0.80 (.75)
African American	-12.75 (5.00)*	-5.16 (1.54)**
Hispanic	-5.57 (3.38)	0.04 (.93)
Asian	2.75 (6.69)	2.05 (1.76)
Other Race/Ethnic Group	-17.81 (5.79)**	-3.44 (1.44)*
Kindergarten SES	-0.18 (0.25)	1.54 (.73)*
Age at Fall Kindergarten Assessment	8.19 (2.93)**	-0.43 (.11)***
Random Effects		
Initial Level Residual Variance	195.66 (121.78)	36.34 (21.87)
Slope Residual Variance	205.53 (121.42)	31.01 (21.44)

Table 41 (cont'd)

Initial Level/Slope Correlation	0.97 (.02)***	0.90 (.06)***
Initial Level/Slope Residual Covariance	195.23 (121.54)	30.32 (21.71)

Note. Values preceded by = indicate that the parameter was fixed to the specified value. Values in parentheses are standard errors. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 42

Fit Statistics for Multiple Group LGM for Eighth Grade Outcomes

Fit Statistics	**Unconstrained Model**	**Constrained Model**
χ^2 / df	110.11 / 70	101.66 / 7
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .01-.04
CFI	0.98	0.98
TLI	0.96	0.97

Note. [A]CI = confidence interval. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index. * $p < .05$; ** $p < .01$; *** $p < .001$

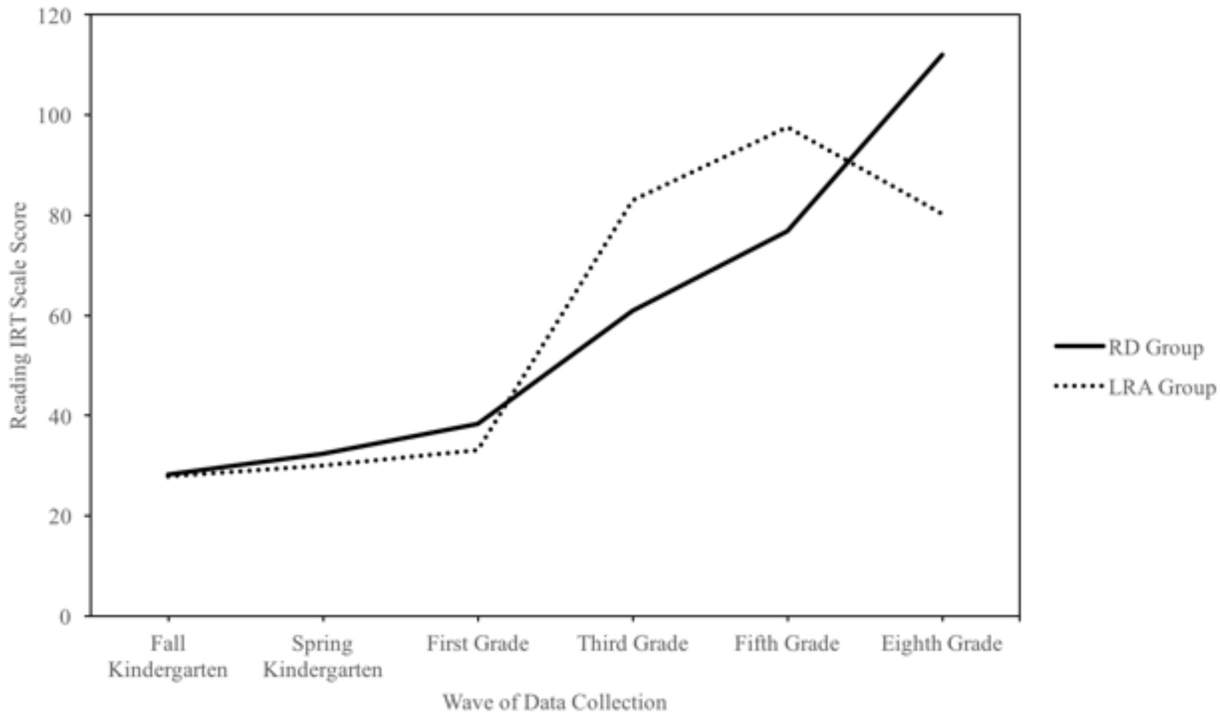

Figure 6. Predicted Reading Growth Trajectories of the RD and LRA Groups

Sensitivity Analyses

Because the two groups were defined using criteria determined by the researcher, it was important to use sensitivity analyses to test the robustness of the above findings. Therefore, whether the above results were maintained when different grouping criteria were used was examined. The base groups that were used in the above analyses were: (1) RD group identified between first and fifth grade, and (2) LRA group identified using weighted average of first, third, and fifth grade T-scores. In each sensitivity analysis, either the RD or the LRA group was replaced with a group defined using different criteria. In other words, the four sensitivity analyses always maintained one of the original RD or LRA groups, while the other group was replaced with a group defined using different criteria. The four sensitivity analyses included: (1) original RD group and an LRA group with children who performed below the 25^{th} percentile at only the third-grade time point; (2) original LRA group and a RD group with children identified between kindergarten and fifth grade; (3) original LRA group and a RD group with children identified between first and fourth grade; and (4) an original LRA group and a RD group with children identified between first and third grade.

For each of these different groupings, a number of models were fit, including the unweighted unconditional latent basis model with residual covariances/regressions, the unweighted conditional latent basis model with residual covariances/regressions, the unweighted multiple group LGM for the initial level, the unweighted multiple group LGM for the shape, and the unweighted multiple group LGM for outcomes in eighth grade. The results for each grouping are described below, and predicted trajectories for each of the sensitivity analyses are shown in Figure 7.

Sensitivity Analysis #1

As described above, the first sensitivity analysis replaced the original LRA group with an LRA group that included children who performed below the 25th percentile on the direct reading assessment at the third-grade time point. The RD group was maintained (i.e., identified between first and fifth grade). The results for each of the different models with this grouping are described below.

Unconditional latent basis model with residual covariances/regressions. The unconditional latent basis model with residual covariances/regressions produced results with a negative slope variance (-3.165, $p = 0.19$), suggesting that this model may not have been the best for the data with the groups defined in the manner. Model fit was slightly worse with these group in comparison to model fit with the original groups. Furthermore, this model produced two negative basis coefficients, which does not seem to be an accurate portrayal of the reading achievement trajectory over time. As a result, there are two possible explanations for these findings: (1) the model with group defined in this manner was not a good fit for the data, or (2) there was something different about the reading achievement trajectory for students in the LRA based on a single third grade achievement score below the 25th percentile as compared to an LRA group using a weighted average T-score below the 25th percentile. Please see Table 43 for the parameter estimates and fit statistics for the sensitivity analysis of the unweighted unconditional model.

Conditional latent basis model with residual covariances/regressions. Despite possible differences in results based on different LRA group for the unconditional model, there were minimal differences in results for the conditional model when considering different definitions of the LRA group. The only noteworthy difference in the two conditional models appeared to be the relationship between the total change (i.e., shape) and kindergarten SES. In the conditional model with the LRA group defined using a weighted average, there was a significant relationship between the total change and kindergarten SES. However, in the model with the LRA group defined using the single third-grade assessment score, the relationship between the total change and SES was not significant. In other words, children with a lower/higher SES did not make significantly less/more progress between kindergarten and eighth grade, such that there was no effect of SES on total change over time. Consequently, defining the LRA group using a single test score produced minimally different results than when the LRA group was defined using a weighted average. Please see Table 44 for the parameter estimates and fit statistics for the sensitivity analysis of the unweighted conditional model.

Multiple group LGM for initial level. The multiple group LGM for the initial level yielded similar results when the LRA group included students who performed below the 25th percentile at only the third-grade time point. Please see Table 45 for fit statistics of the constrained and unconstrained models. Results indicated no significant group differences in the initial starting levels ($\Delta\chi^2 = 1.48$, $\Delta df = 1$) at the .05 level. Results of the Wald test also supported the finding of a non-significant difference in the initial level of reading achievement between the two groups ($t = 1.50$, df = 1, $p = .22$). Please note that the unconstrained model produced results with a negative slope variance in the LRA group (-0.02, $p = .99$), suggesting that this model may not have been the best for the data with the groups defined in the manner. Although this may not have been the most appropriate model for these data, the results suggest that the RD group and the LRA group had similar levels of reading achievement in the fall of kindergarten, even when the LRA group was defined without using a weighted average T-score below the 25th percentile.

Multiple group LGM for shape. The multiple group LGM for the shape yielded similar results when the LRA group included students who performed below the 25th percentile at only the third-grade time point. Results indicated significant group differences in the shape of reading achievement over time ($\Delta\chi^2 = 85.80$, $\Delta df = 4$) at the .05 level. Please see Table 46 for fit statistics of the constrained and unconstrained models. Of note, the unconstrained model produced results with a negative slope variance in the LRA group (-0.02, $p = .99$), and the constrained model produced results with a negative slope variance in the RD group (-8.20, $p = .90$). These results suggest that this model may not have been the best for the data with the groups defined in the manner. Although this may not have been the most appropriate model for these data, it was still important to examine whether group differences existed in this grouping pair. In general, these results suggest that the RD group and the LRA group had different shapes of reading achievement, even when the LRA group was defined without using a weighted average T-score below the 25th percentile.

Multiple group LGM for outcomes in eighth grade. The multiple group LGM for eighth grade as the initial level yielded conflicting results when the LRA group included students who performed below the 25th percentile at only the third-grade time point. Although the chi-square difference test could not be computed due to a negative difference value, a Wald test revealed no significant group differences in the reading achievement scores at the end of eighth grade ($t = 2.42$, df = 1, $p = .12$). A Wald test (i.e., independent samples t-test) that compared the mean eighth-grade reading achievement scores revealed marginally significant group differences ($t = 4.01$, df = 1, $p = .04$). Here, the estimated mean IRT scale score for the RD group was 141.81 and for the LRA group was 137.62, which were quite similar to the means of the original groups. Also of note, the unconstrained model produced results with a negative slope and initial level variance in both the RD group (initial level: -22.45, $p = .16$; slope: -18.96, $p = .08$) and in the LRA group (initial level: -1.27, $p = .81$; slope: -0.02, $p = .99$), and the constrained model produced a negative initial level variance in the LRA group (-0.18, $p = .98$). Further, in the graph of the predicted reading growth trajectories (see Figure 7), it appears that there was a significant difference in the eighth-grade reading outcomes between the two groups. Thus, it is possible that the conflicting findings occurred because of poor model fit and inappropriate basis coefficients as well as slight differences in the reading achievement trajectory and relationship between the covariates when the LRA group was defined using a single achievement score below the 25th percentile. Please see Table 47 for fit statistics of the constrained and unconstrained models.

Table 43

Parameter Estimates and Fit Statistics for Sensitivity Analyses of Unweighted Unconditional Latent Basis Models

Parameter	Original Groups (RD 1st-5th and LRA Weighted Average) n = 1080	Sensitivity Analysis #1 (RD 1st-5th and LRA 3rd grade only) n = 1010	Sensitivity Analysis #2 (RD K-5th and LRA Weighted Average) n = 1120	Sensitivity Analysis #3 (RD 1st-4th and LRA Weighted Average) n = 1070	Sensitivity Analysis #4 (RD 1st-3rd and LRA Weighted Average) n = 1030
Fixed Effects					
Mean Initial Level	28.32 (.14)***	28.63 (.15)***	28.36 (.14)***	28.31 (.14)***	28.27 (.14)***
Mean Slope	77.40 (25.66)**	16.13 (7.39)*	80.26 (27.84)**	76.82 (25.56)**	70.41 (24.31)**
Basis Coefficients	=0	=0	=0	=0	=0
	0.03 (.02)	-0.15 (.15)	0.03 (.02)	0.03 (.02)	0.04 (.03)
	0.08 (.02)*	-0.01 (.13)	0.07 (.03)*	0.08 (.04)*	0.08 (.04)*
	0.59 (.20)**	1.97 (.87)*	0.55 (.19)**	0.60 (.20)**	0.65 (.23)**
	0.87 (.25)***	1.36 (.48)**	0.80 (.23)**	0.88 (.25)***	0.92 (.26)***
	=1	=1	=1	=1	=1
Random Effects					
Initial Level Variance	10.26 (1.10)***	10.15 (2.03)***	10.10 (1.07)***	10.21 (1.10)***	10.09 (1.14)***
Slope Variance	158.64 (117.33)	-3.17 (2.42)	178.43 (137.76)	155.24 (115.69)	126.53 (100.40)
Initial Level/Slope Correlation	0.02 (.08)	-------------------[B]	0.03 (.08)	0.77 (.80)	-0.01 (.09)
Initial Level/Slope Covariance	0.64 (3.26)	-2.62 (2.03)***	1.29 (3.59)	0.96 (3.27)	-0.18 (3.03)
Fit Statistics					
χ^2 / df	10.52 / 7	34.44 / 7	11.872 / 7	10.43 / 7	11.47 / 7
RMSEA / CI[A]	0.02 / .00-.05	0.06 / .04-.08	0.03 / .00-.05	0.02 / .00-.05	0.03 / .00-.05
CFI	0.998	0.98	0.997	0.998	0.997
TLI	0.995	0.97	0.99	0.995	0.993

Note. [A]CI = confidence interval. [B]Not Available. Values preceded by = indicate that the parameter was fixed to the specified value. Values in parentheses are standard errors. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 44

Parameter Estimates and Fit Statistics for Sensitivity Analyses of Unweighted Conditional Latent Basis Models

Parameter	Original Groups (RD 1st-5th and LRA Weighted Average) *n* = 1080	Sensitivity Analysis #1 (RD 1st-5th and LRA 3rd Grade Only) *n* = 1010	Sensitivity Analysis #2 (RD K-5th and LRA Weighted Average) *n* = 1120	Sensitivity Analysis #3 (RD 1st-4th and LRA Weighted Average) *n* = 1070	Sensitivity Analysis #4 (RD 1st-3rd and LRA Weighted Average) *n* = 1030
Fixed Effects					
Mean (Intercept) Initial Level	12.78 (1.96)***	13.19 (2.34)***	13.45 (1.87)***	12.81 (1.97)***	13.22 (2.02)***
Mean (Intercept) Slope	125.04 (30.19)***	68.70 (28.05)*	125.23 (31.21)***	125.12 (30.29)***	127.44 (31.20)***
Basis Coefficients	=0	=0	=0	=0	=0
	0.03 (.02)	0.06 (.04)	0.04 (0.02)	0.03 (.02)	0.03 (.02)
	0.08 (.03)*	0.12 (.04)**	0.07 (.03)*	0.08 (.03)*	0.07 (.03)*
	0.52 (.14)***	0.83 (.31)**	0.51 (.14)***	0.52 (0.14)***	0.53 (.14)***
	0.75 (.16)***	0.86 (.16)***	0.71 (.15)***	0.75 (.16)***	0.75 (.16)***
	=1	=1		=1	=1
Initial Level on Covariates					
Female	0.74 (.24)**	0.61 (.29)*	0.75 (.24)**	0.77 (.24)**	0.77 (.25)**
African American	-0.28 (.32)	0.21 (.41)	-0.25 (.31)	-0.31 (.32)	-0.22 (.33)
Hispanic	-1.26 (.33)***	-1.85 (.37)***	-1.16 (.33)***	-1.25 (.33)***	-1.12 (.34)**
Asian	-1.72 (.74)*	-2.74 (.78)***	-1.67 (.75)*	-1.71 (.74)*	-1.45 (.73)
Other Race/Ethnic Group	-1.68 (.40)***	-1.25 (.49)*	-1.50 (.39)***	-1.68 (.40)***	-1.58 (.41)***
Kindergarten SES	1.04 (.18)***	1.27 (.21)***	1.03 (.18)***	1.04 (.18)***	1.01 (.19)***
Age at Fall Kindergarten	0.24 (.03)***	0.24 (.03)***	0.23 (.03)***	0.24 (.03)***	0.23 (.03)***
Shape on Covariates					
Female	2.00 (1.27)	1.14 (1.16)	2.06 (1.29)	1.87 (1.26)	1.85 (1.29)
African American	-11.57 (3.21)***	-7.52 (2.86)**	-11.37 (3.24)***	-11.50 (3.21)***	-11.40 (3.25)***
Hispanic	-3.04 (1.86)	-2.50 (1.46)	-2.49 (1.84)	-3.06 (1.86)	-2.99 (3.25)
Asian	0.47 (2.80)	0.48 (2.15)	1.12 (2.86)	0.44 (2.79)	0.32 (2.81)
Other Race/Ethnic Group	-9.91 (3.13)**	-5.43 (2.34)*	-8.72 (3.06)**	-9.90 (3.13)**	-9.22 (3.01)**

Table 44 (cont'd)

Kindergarten SES	4.45 (1.58)**	2.31 (1.31)	4.72 (1.68)**	4.41 (1.57)**	4.04 (1.53)**
Age at Fall Kindergarten	-0.49 (.14)***	-0.32 (.11)**	-0.50 (.14)***	-0.50 (.14)***	-0.56 (.15)***
Random Effects					
Initial Level Residual Variance	8.11 (.97)***	11.39 (1.78)***	8.15 (.93)***	8.03 (.96)***	8.04 (.98)***
Slope Residual Variance	161.50 (99.38)	25.79 (48.00)	171.07 (110.41)	159.39 (99.01)	152.44 (97.08)
Initial Level/Slope Correlation	-0.02 (.08)	-0.25 (.24)	-0.01 (.08)	-0.01 (.08)	-0.03 (.08)
Initial Level/Slope Residual Covariance	-2.23 (2.88)	-6.65 (2.83)	-1.40 (2.95)	-1.88 (2.88)	-1.87 (2.93)
Fit Statistics					
χ^2 / df	61.41 / 35	108.36 / 35	57.70 / 35	60.80 / 35	57.38 / 35
RMSEA / CI[A]	0.03 / .02-.04	0.05 / .04-.06	0.02 / .01-.04	0.03 / .02-.04	0.03 / .01-.04
CFI	0.99	0.97	0.99	0.99	0.99
TLI	0.98	0.94	0.98	0.98	0.98

Note. [A]CI = confidence interval. Values preceded by = indicate that the parameter was fixed to the specified value. Values in parentheses are standard errors. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 45

Sensitivity Analysis #1: Parameter Estimates for Unconstrained and Constrained Initial Level Model

Fit Statistics	Original Groups	Sensitivity Group #1
	Unconstrained Model	
χ^2 / df	101.87 / 70	119.05 / 70
RMSEA / CI[A]	0.03 / .02-.04	0.04 / .03-.05
CFI	0.98	0.98
TLI	0.97	0.96
	Constrained Model	
χ^2 / df	104.12 / 71	120.47 / 71
RMSEA / CI[A]	0.03 / .02-.04	0.04 / .03-.05
CFI	0.98	0.98
TLI	0.97	0.96

Note. [A]CI = confidence interval. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index.

Table 46

Sensitivity Analysis #1: Parameter Estimates for Unconstrained and Constrained Basis Coefficients Model

Fit Statistics	Original Groups	Sensitivity Group #1
	Unconstrained Model	
χ^2 / df	101.87 / 70	119.05 / 70
RMSEA / CI[A]	0.03 / .02-.04	0.04 / .03-.05
CFI	0.98	0.98
TLI	0.97	0.96
	Constrained Model	
χ^2 / df	115.13 / 74	142.12 / 74
RMSEA / CI[A]	0.03 / .02-.04	0.04 / .03-.05
CFI	0.98	0.97
TLI	0.97	0.95

Note. [A]CI = confidence interval. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index.

Table 47

Sensitivity Analysis #1: Parameter Estimates for Unconstrained and Constrained Eighth Grade Initial Level Model

Fit Statistics	Original Groups	Sensitivity Group #1
	Unconstrained Model	
χ^2 / df	110.11 / 70	127.45 / 70
RMSEA / CI[A]	0.03 / .02-.04	0.04 / .03-.05
CFI	0.98	0.97
TLI	0.96	0.96
	Constrained Model	
χ^2 / df	101.66 / 71	120.416 / 71
RMSEA / CI[A]	0.03 / .01-.04	0.04 / .03-.05
CFI	0.98	0.98
TLI	0.97	0.96

Note. [A]CI = confidence interval. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index.

Sensitivity Analysis #2

As described above, the second sensitivity analysis replaced the original RD group with a RD group that included children who were identified to received special education services due to RD between kindergarten and fifth grade. The original LRA group was maintained (i.e., weighted average below 25th percentile). The results for each of the different models with this grouping are described below.

Unconditional latent basis model with residual covariances/regressions. When kindergarten students were included in the RD group, the results for the unconditional latent basis model with residual covariances/regressions were extremely similar to the results when the kindergarten students were not included in the RD group (see Table 43). Therefore, it is unlikely that the overall reading growth trajectory would have looked different had the RD group included children identified in kindergarten.

Conditional latent basis model with residual covariances/regressions. The results for the conditional latent basis model with residual covariances/regressions were very similar when the RD group both included and excluded kindergarten students identified as having a RD (see Table 44). Thus, the relationship between reading achievement growth and the covariates did not appear to change when RD-identified kindergarten students were included in the RD group.

Multiple group LGM for initial level. The multiple group LGM for the initial level yielded similar results when the RD group included students who were identified to receive special education services due to a RD in kindergarten. Please see Table 48 for fit statistics of the unconstrained and constrained models. Results indicated no significant group differences in the initial starting levels ($\Delta\chi^2 = 0.96$, $\Delta df = 1$) at the .05 level. Results of the Wald test also supported the finding of a non-significant difference in the initial level of reading achievement between the two groups ($t = 0.92$, df = 1, $p = .34$). Therefore, these results suggest that the RD group and the LRA group had similar levels of reading achievement in the fall of kindergarten, even when the RD group included children who were identified in kindergarten.

Multiple group LGM for shape. The multiple group LGM for the shape yielded similar results when the when the RD group included students who were identified to received special education services due to a RD in kindergarten. Results indicated significant group differences in the shape of reading achievement over time ($\Delta\chi^2 = 12.41$, $\Delta df = 4$) at the .05 level. Please see Table 49 for fit statistics of the constrained and unconstrained models. Of importance, the constrained model produced results with a negative slope variance in the RD group (-10.86, p = .68). Therefore, these results suggest that this model may not have been the best for the data with the groups defined in the manner. Although this may not have been the most appropriate model for these data, the results suggest that the RD group and the LRA group had different patterns, or shapes of reading achievement over time, even when the RD group included children who were identified with RD in kindergarten.

Multiple group LGM for outcomes in eighth grade. The multiple group LGM for eighth grade as the initial level yielded similar results when the RD group included students who were identified to received special education services due to a RD in kindergarten. Please see Table 50 for fit statistics of the constrained and unconstrained models. Although the chi-square difference test could not be computed due to a negative difference value, a Wald test revealed significant group differences in the reading achievement scores at the end of eighth grade (t = 6.99, df = 1, p < .01). Further, a Wald test (i.e., independent samples t-test) that compared the mean eighth-grade reading achievement scores revealed similar results (t = 4.76, df = 1, p < .05). As in the original groups, the RD group had a significantly higher mean IRT scale score of 140.99 compared to the LRA group's mean IRT scale score of 136.92. Also of note, the unconstrained model produced results with a negative slope and initial level variance in the RD group (initial level: -20.26, p = .30; slope: -19.19, p = .20). Even though this may not have been the most appropriate model for these data, these results suggest that the RD group and the LRA group had different levels of reading achievement in the spring of eighth grade, even when the RD group included children who were identified in kindergarten.

Table 48

Sensitivity Analysis #2: Parameter Estimates for Unconstrained and Constrained Initial Level Model

Fit Statistics	Original Groups	Sensitivity Group #2
Unconstrained Model		
χ^2 / df	101.87 / 70	99.45 / 70
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .01-.04
CFI	0.98	0.98
TLI	0.97	0.98
Constrained Model		
χ^2 / df	104.12 / 71	100.31 / 71
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .01-.04
CFI	0.98	0.98
TLI	0.97	0.98

Note. [A]CI = confidence interval. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index.

Table 49

Sensitivity Analysis #2: Parameter Estimates for Unconstrained and Constrained Basis Coefficients Model

Fit Statistics	Original Groups	Sensitivity Group #2
Unconstrained Model		
χ^2 / df	101.87 / 70	99.45 / 70
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .01-.04
CFI	0.98	0.98
TLI	0.97	0.98
Constrained Model		
χ^2 / df	115.13 / 74	110.57 / 74
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .02-.04
CFI	0.98	0.98
TLI	0.97	0.97

Note. [A]CI = confidence interval. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index.

Table 50

Sensitivity Analysis #2: Parameter Estimates for Unconstrained and Constrained Eighth Grade Initial Level Model

Fit Statistics	Original Groups	Sensitivity Group #2
	Unconstrained Model	
χ^2 / df	110.11 / 70	101.18 / 70
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .02-.04
CFI	0.98	0.98
TLI	0.96	0.97
	Constrained Model	
χ^2 / df	101.66 / 71	99.70 / 71
RMSEA / CI[A]	0.03 / .01-.04	0.03 / .01-.04
CFI	0.98	0.99
TLI	0.97	0.98

Note. [A]CI = confidence interval. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index.

Sensitivity Analysis #3

As described above, the third sensitivity analysis replaced the original RD group with a RD group that included children who were identified to received special education services due to RD between first and fourth grade. The original LRA group was maintained (i.e., weighted average below 25th percentile). The results for each of the different models with this grouping are described below.

Unconditional latent basis model with residual covariances/regressions. When fifth-grade students were excluded from the RD group, the results for the unconditional latent basis model with residual covariances/regressions were extremely similar to the results when the fifth-grade students were included in the RD group (see Table 43). Therefore, it is unlikely that the overall reading growth trajectory would have looked different had the RD group excluded children identified in fifth grade.

Conditional latent basis model with residual covariances/regressions. The results for the conditional latent basis model with residual covariances/regressions were very similar when the RD group both included and excluded fifth-grade students identified as having a RD (see Table 44). Thus, the relationship between reading achievement growth and the covariates did not appear to change when RD-identified fifth-grade students were excluded from the RD group.

Multiple group LGM for initial level. The multiple group LGM for the initial level yielded similar results when the RD group excluded students who were identified to receive special education services due to a RD in fifth grade. Please see Table 51 for fit statistics of the constrained and unconstrained models. Results indicated no significant group differences in the initial starting levels ($\Delta\chi^2 = 2.29$, $\Delta df = 1$) at the .05 level. Results of the Wald test also supported the finding of a non-significant difference in the initial level of reading achievement between the two groups ($t = 2.20$, df = 1, $p = .14$). Therefore, these results suggest that the RD group and the LRA group had similar levels of reading achievement in the fall of kindergarten, even when the RD group excluded children who were identified in fifth grade.

Multiple group LGM for shape. The multiple group LGM for the shape yielded similar results when the RD group excluded students who were identified to receive special education services due to a RD in fifth grade. Results indicated significant group differences in the shape of reading achievement over time ($\Delta\chi^2 = 12.63$, $\Delta df = 4$) at the .05 level. Please see Table 52 for fit statistics of the constrained and unconstrained models. Of importance, the constrained model produced results with a negative slope variance in the RD group (-1.56, $p = .96$). Therefore, these results suggest that this model may not have been the best for the data with the groups defined in the manner. Although this may not have been the most appropriate model for these data, the results suggest that the RD group and the LRA group had different patterns, or shapes of reading achievement over time, even when the RD group excluded children who were identified with RD in fifth grade.

Multiple group LGM for outcomes in eighth grade. The multiple group LGM for eighth grade as the initial level yielded similar results RD group excluded students who were identified to receive special education services due to a RD in fifth grade. Please see Table 53 for fit statistics of the constrained and unconstrained models. Although the chi-square difference test could not be computed due to a negative difference value, a Wald test revealed significant group differences in the reading achievement scores at the end of eighth grade (t = 4.47, df = 1, p < .05). Further, a Wald test (i.e., independent samples t-test) that compared the mean eighth-grade reading achievement scores revealed similar results (t = 6.51, df = 1, p < .05). As in the original groups, the RD group had a significantly higher mean IRT scale score of 142.04 compared to the LRA group's mean IRT scale score of 136.92. Also, please note that the unconstrained model produced results with a negative slope and initial level variance in the RD group (initial level variance: -22.52, p = .16; slope variance: -19.28, p = .08). Even though this may not have been the most appropriate model for these data, these results suggest that the RD group and the LRA group had different levels of reading achievement in the spring of eighth grade, even when the RD group included children who were identified to receive special education services due to a RD in fifth grade.

Table 51

Sensitivity Analysis #3: Parameter Estimates for Unconstrained and Constrained Initial Level Model

Fit Statistics	Original Groups	Sensitivity Group #3
	Unconstrained Model	
χ^2 / df	101.87 / 70	104.80 / 70
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .02-.04
CFI	0.98	0.98
TLI	0.97	0.97
	Constrained Model	
χ^2 / df	104.12 / 71	107.27 / 71
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .02-.04
CFI	0.98	0.98
TLI	0.97	0.97

Note. [A]CI = confidence interval. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index.

Table 52

Sensitivity Analysis #3: Parameter Estimates for Unconstrained and Constrained Basis Coefficients Model

Fit Statistics	Original Groups	Sensitivity Group #3
	Unconstrained Model	
χ^2 / df	101.87 / 70	104.80 / 70
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .02-.04
CFI	0.98	0.98
TLI	0.97	0.97
	Constrained Model	
χ^2 / df	115.13 / 74	117.25 / 74
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .02-.04
CFI	0.98	0.98
TLI	0.97	0.96

Note. [A]CI = confidence interval. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index.

Table 53

Sensitivity Analysis #3: Parameter Estimates for Unconstrained and Constrained Eighth Grade Initial Level Model

Fit Statistics	Original Groups	Sensitivity Group #3
	Unconstrained Model	
χ^2 / df	110.11 / 70	112.25 / 70
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .02-.05
CFI	0.98	0.98
TLI	0.96	0.96
	Constrained Model	
χ^2 / df	101.66 / 71	104.60 / 71
RMSEA / CI[A]	0.03 / .01-.04	0.03 / .02-.04
CFI	0.98	0.98
TLI	0.97	0.97

Note. [A]CI = confidence interval. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index.

Sensitivity Analysis #4

As described above, the fourth sensitivity analysis replaced the original RD group with a RD group that included children who were identified to received special education services due to RD between first and third grade. The original LRA group was maintained (i.e., weighted average below 25th percentile). The results for each of the different models with this grouping are described below.

Unconditional latent basis model with residual covariances/regressions. When fourth and fifth grade students were excluded from the RD group, the results for the unconditional latent basis model with residual covariances/regressions were extremely similar to the results when fourth- and fifth-grade students were included in the RD group (see Table 43). Therefore, it is unlikely that the overall reading growth trajectory would have looked different had the RD group excluded children identified in fourth and fifth grade.

Conditional latent basis model with residual covariances/regressions. There appeared to be only one difference in results between the models when the RD group either included or excluded fourth- and fifth-grade students (see Table 44). Specifically, in the model where the fourth and fifth grade students were included in the RD group (i.e., original analyses), there was a significant relationship between being an Asian student and the initial level of reading achievement, such that Asian students had lower initial levels of reading achievement than White students. Nonetheless, when fourth- and fifth-grade students were excluded from the RD group, being an Asian student did not significantly predict a lower initial achievement level. There were no other differences in the results for the conditional latent basis model with residual covariances/regressions when fourth- and fifth-grade students were excluded. Thus, the relationship between reading achievement growth and the covariates did not appear to change when RD-identified fourth- and fifth-grade students were excluded from the RD group, except for the relationship between the initial level and Asian students.

Multiple group LGM for initial level. The multiple group LGM for the initial level yielded similar results when the RD group excluded students who were identified to receive special education services due to a RD in fourth and fifth grade. Please see Table 54 for fit statistics of the constrained and unconstrained models. Results indicated no significant group differences in the initial starting levels ($\Delta\chi^2 = 1.60$, $\Delta df = 1$) at the .05 level. Results of the Wald test also support the finding of a non-significant difference in the initial level of reading achievement between the two groups ($t = 1.53$, df = 1, $p = .22$). Therefore, these results suggest that the RD group and the LRA group had similar levels of reading achievement in the fall of kindergarten, even when the RD group excluded children who were identified in fourth and fifth grade.

Multiple group LGM for shape. The multiple group LGM for the shape yielded similar results when the RD group excluded students who were identified to receive special education services due to a RD in fourth and fifth grade. Results indicated significant group differences in the shape of reading achievement over time ($\Delta\chi^2 = 11.10$, $\Delta df = 4$) at the .05 level. Please see Table 55 for fit statistics of the constrained and unconstrained models. Of importance, the constrained model produced results with a negative slope/initial level covariance in the RD group (-4.92, $p = .16$), which occurred due to a non-positive definite covariance matrix. Therefore, these results suggest that this model may not have been the best for the data with the groups defined in the manner. Although this may not have been the most appropriate model for these data, the results suggest that the RD group and the LRA group had different patterns, or shapes of reading achievement over time, even when the RD group excluded children who were identified to receive special education services with RD in fourth and fifth grade.

Multiple group LGM for outcomes in eighth grade. The multiple group LGM for eighth grade as the initial level yielded conflicting results when the RD group excluded students who were identified to receive special education services due to a RD in fourth and fifth grade. Please see Table 56 for fit statistics of the constrained and unconstrained models. Results indicated no significant group differences in reading achievement outcomes in eighth grade ($\Delta\chi^2 = 0.20$, $\Delta df = 1$) at the .05 level. In support of this finding, a Wald test revealed no significant group differences in the reading achievement scores at the end of eighth grade ($t = 0.18$, df = 1, $p = .68$). Similarly, a Wald test (i.e., independent samples t-test) that compared the mean eighth-grade reading achievement scores revealed similar results ($t = 3.96$, df = 1, $p = .05$). The mean IRT scale score of the RD and LRA groups were 141.21 and 136.92, respectively. Although these means are similar to those of the original groups, this model with the RD group excluding children identified after third grade indicated that there were not differences between the two groups' reading achievement at the end of eighth grade. Perhaps, excluding children who were identified in fourth and fifth grade removed some of the inter- and intra-individual variation in reading scores to make the two groups look more similar. In other words, those identified in later grades may have been those with the higher reading scores in eighth grade, and by removing them from the sample, the variation they contributed to the group was removed as well. Also of importance, this model ran without any errors related to negative slope and initial level variance, so it may be that this grouping permits for the drawing of more appropriate conclusions compared to the model with the original groupings. Further consideration and discussion about a RD group that excludes students identified with RD in fourth and fifth grade is provided below.

Table 54

Sensitivity Analysis #4: Parameter Estimates for Unconstrained and Constrained Initial Level Model

Fit Statistics	**Original Groups**	**Sensitivity Group #4**
	Unconstrained Model	
χ^2 / df	101.87 / 70	103.04 / 70
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .02-.04
CFI	0.98	0.98
TLI	0.97	0.97
	Constrained Model	
χ^2 / df	104.12 / 71	104.68 / 71
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .02-.04
CFI	0.98	0.98
TLI	0.97	0.97

Note. [A]CI = confidence interval. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index.

Table 55

Sensitivity Analysis #4: Parameter Estimates for Unconstrained and Constrained Basis Coefficients Model

Fit Statistics	**Original Groups**	**Sensitivity Group #4**
	Unconstrained Model	
χ^2 / df	101.87 / 70	103.04 / 70
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .02-.04
CFI	0.98	0.98
TLI	0.97	0.97
	Constrained Model	
χ^2 / df	115.13 / 74	114.48 / 74
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .02-.04
CFI	0.98	0.98
TLI	0.97	0.96

Note. [A]CI = confidence interval. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index.

Table 56

Sensitivity Analysis #4: Parameter Estimates for Unconstrained and Constrained Eighth Grade Initial Level Model

Fit Statistics	**Original Groups**	**Sensitivity Group #4**
	Unconstrained Model	
χ^2 / df	110.11 / 70	103.04 / 70
RMSEA / CI[A]	0.03 / .02-.04	0.03 / .02-.04
CFI	0.98	0.98
TLI	0.96	0.97
	Constrained Model	
χ^2 / df	101.66 / 71	102.76 / 71
RMSEA / CI[A]	0.03 / .01-.04	0.03 / .02-.04
CFI	0.98	0.98
TLI	0.97	0.97

Note. [A]CI = confidence interval. χ^2= maximum likelihood chi square fit statistic, RMSEA = root mean square error of approximation, CFI = comparative fit index.

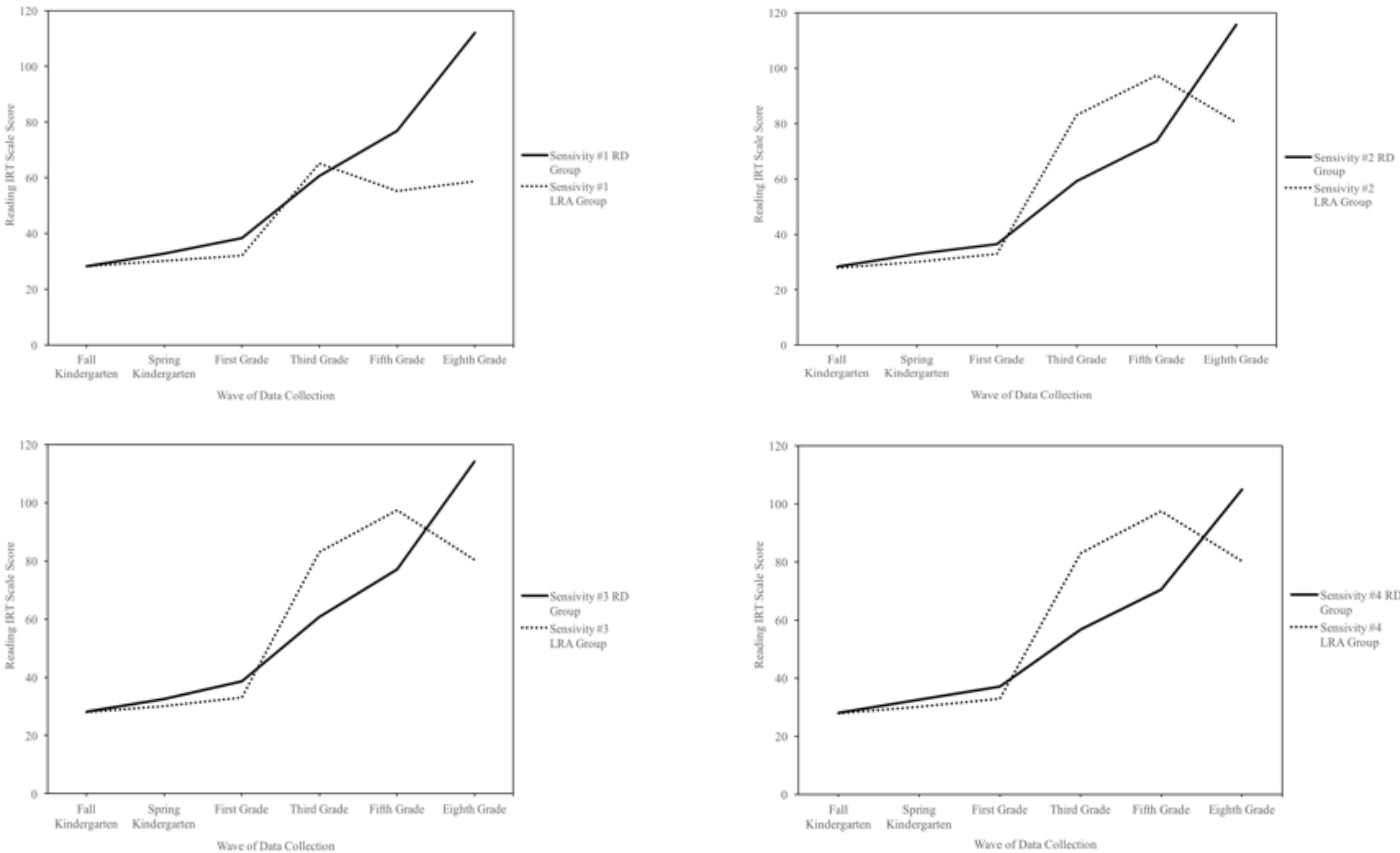

Figure 7. Predicted Reading Growth Trajectories of the Sensitivity Analyses for the RD and LRA Groups. Top left: Sensitivity #1 RD 1st-5th and LRA 3rd grade only below 25th percentile; top right: Sensitivity #2 RD K-5th and LRA weighted average below 25th percentile; bottom left: Sensitivity #3 RD 1st-4th and LRA weighted average below 25th percentile; bottom right: Sensitivity #4 RD 1st-3rd and LRA weighted average below 25th percentile.

Summary

First, unweighted and weighted descriptive statistics were examined. The unweighted descriptive statistics revealed that the full sample was composed of lower SES, racially/ethnically diverse students around five and a half years of age. When comparing the RD group to the LRA group, the RD group had a higher proportion of boys than did the LRA group, while the LRA group had a lower SES and was more racially/ethnically diverse than the RD group. The RD group also outperformed the LRA group at the third- and eighth-grade time points on the direct reading assessment.

Results of the weighted descriptive statistics were relatively similar to the unweighted results. When the weights were used, the RD group became 8% more racially/ethnically diverse, while the LRA group stayed the same. All other weighted descriptive statistics were quite similar to their unweighted counterparts. However, when the weights were applied, more statistically significant differences were found between the RD and LRA group's reading IRT scale scores, but it did not appear that all of these differences were clinically significant. Therefore, examination of the descriptive statistics indicated that the weighted and unweighted samples were fairly similar. Still, it should be noted that analyses performed using unweighted data does not generalize to the 1998-1999 kindergarten class.

Second, a logistic regression was conducted to determine whether the covariates predicted group membership. Unweighted results indicated that all of the independent variables predicted membership in the RD group. Specifically, boys were more likely to be identified as members of the RD group, while African American, Hispanic, Asian, and students from other diverse racial/ethnic group were significantly less likely to be identified for the RD group. Older students and students with higher SES were more likely to be identified for the RD group.

Similar results were found when using weighted data, except that the gender dummy variable did not significantly predict group membership. Although gender did not predict group membership, it was still used as a covariate in the LGM model, given that gender plays an important role in reading development.

Next, unweighted and weighted unconditional and conditional LGMs were fit to the data. The weighted data produced unreliable results due to negative slope and intercept variances as well as implausible basis coefficient estimates. As a result, only the unweighted data were analyzed in depth. For the unweighted unconditional model, a latent basis model with residual covariances/regressions was the best model to capture the longitudinal trajectory of the full sample's reading achievement over time. The basis coefficients in this unconditional latent basis model suggested that 59% of the growth occurred between the fall of kindergarten and the spring of third grade, and 87% of the growth occurred between the fall of kindergarten and the spring of fifth grade. In other words, over half of the growth in reading achievement occurred between the spring of first grade and the spring of third grade, with another 28% of growth occurring between the spring of third grade and the spring of fifth grade, and the remaining 13% occurring between the spring fifth grade and the spring of eighth grade.

The results of unweighted conditional latent basis model reveled important relationships among initial levels of reading achievement, total growth, and the covariates. More specifically, Hispanic children, Asian children, other racially/ethnic diverse children, and children with lower SES had significantly lower reading achievement at the fall of kindergarten. There were no significant differences between African American children and White children's reading achievement at the start of kindergarten, and older children had significant higher reading achievement at the start of kindergarten. In terms of total growth over time, African American children, other racially/ethnic diverse children, and younger children made significantly less gains in reading achievement over time, whereas children with higher SES made significantly greater progress over time. Girls, Hispanic children, and Asian children did not significantly differ from the boys and White children, respectively in terms of total growth in reading achievement over time.

Next, to examine group differences, three multiple group LGMs were carried out using the unweighted data. The multiple group LGMs revealed that the two groups did not differ in terms of initial starting levels of reading achievement. The multiple group LGM did reveal, however, that the two groups differed in terms of the shape of total change over time as well as in terms of reading achievement outcomes at the end of eighth grade. Specifically, at the end of eighth grade, the RD group outperformed the LRA group.

Finally, sensitivity analyses with different groups were conducted in order to test the robustness of the original findings. By and large, the results of the sensitivity analyses were consistent with the results based on the original groups (i.e., LRA weighted average below 25th percentile; RD 1st-5th grade). However, some differences were found. In sensitivity analysis #1 with the LRA group defined using a single IRT scale score below the 25th percentile, the unconditional and conditional latent basis models were not the best fit for the data, which may explain some of the conflicting results with this grouping. Specifically, no group differences were found in eighth-grade reading achievement outcomes. Thus, it is possible that there was something different about an LRA group defined using a single score that caused results that were slightly different from those using the original grouping. Results of sensitivity analysis #2, in which the RD group included children identified in kindergarten through fifth grade, were the same as those when the RD excluded children diagnosed with RD in kindergarten. Similarly, results of sensitivity analysis #3, in which the RD group included children identified in first through fourth grade, were the same as those when the RD group included children diagnosed with RD in fifth grade. Lastly, results of sensitivity analysis #4, in which the RD group included children first through third grade, were the same as those when the RD group included children diagnosed in fourth and fifth grade, except that no differences between the LRA and RD groups' reading achievement were found at the end of eighth grade. Given that two of the sensitivity analyses led to conflicting results at the eighth-grade time point, careful consideration and discussion is warranted as to why differing results were found.

CHAPTER 5
DISCUSSION

Researchers continue to study the reading development and reading achievement of children who struggle to learn to read. Around 7% of children who have difficulty with reading will be diagnosed with RD, and many of these children will receive special education services in school. However, there are also students who struggle to read, but do not receive a diagnosis of RD or subsequent special education services. These children are often referred to as students with LRA.

Understanding the differences between children with diagnosed RD and with LRA has been of focus in many areas of research. First, researchers have studied whether children with RD and LRA differ in terms of their cognitive abilities and neuropsychological functioning. Generally, this research indicates that the two groups do not differ across such domains. Furthermore, part of this research has determined that both children with RD and LRA demonstrate deficits in phonological awareness, and that reading ability exists on a continuum, such that those with RD are not a separate entity from those with less severe reading problems. Second, researchers have also examined how students with reading problems respond to intervention. From this research, researchers have identified students whose reading problems are easily remediated and students whose reading problems are more difficult to remediate. Even though some students' reading problems may be difficult to remediate, they are still able to make modest improvements to their reading skills with intensive, high-quality, and evidence-based reading interventions. Third, researchers have studied the longitudinal trends in reading achievement and have recognized two contrasting theories on reading achievement trajectories. One of these theories is called the Matthew effect, which suggests that children with the lowest reading achievement will get worse over time, and those with the highest reading skills will make the most progress. The other theory is called the compensatory trajectory of reading development, in which children with initially lower reading skills grow at a faster rate and catch up to their peers with initially higher reading skills. For both the Matthew effect and the compensatory trajectory of reading development, there is evidence that regardless of the rate of progress, children with RD never catch up to their same-age peers. Research on the different patterns of reading growth also allows for the consideration of at what point in time do children experience the most growth in their reading skills. Generally, research suggests that children make the most gains in reading before first grade, and after first grade, growth slows. Thus, when examining the results of the current study, which utilized only a sample of poor readers, it is important to think about how the participants' patterns of reading growth fit within the literature that compares children without reading problems to children with reading problems.

Lastly, and most closely related to the current study, researchers have studied differences in the longitudinal reading achievement trajectories between children with RD and LRA. However, the present literature review identified only two studies that allowed for the comparison of initial reading achievement and rates of growth over time. Both of these studies found differences in neither initial reading achievement scores nor rates of growth over time between the two groups. These studies, however, did not include information about special education eligibility, so whether students received special education services was unknown. This study extended the research, providing additional insight about differences in the reading achievement trajectories between students identified to receive special education services for RD and non-identified students with LRA from kindergarten through eighth grade.

The current longitudinal study utilized LGM modeling to both examine the reading achievement growth trajectories in a sample of children with reading difficulties and compare the trajectories of a special education-identified RD group and a non-identified LRA group. Specifically, differences in initial reading achievement at the start of kindergarten, patterns of growth over time, and reading achievement outcomes at the end of eighth grade were examined. Before conducting the latent growth analyses, a logistic regression was conducted to determine whether the covariates predicted group membership. Logistic regression hypotheses were supported such that all of the covariates significantly predicted group membership in the unweighted model, and only gender did not predict group membership in the weighted model. Next, a latent basis model with residual covariances/regression was identified as the best fitting unconditional model. Then, a conditional latent basis model with residual covariances/regression was fit to the data in order to determine the relationship between the covariates and growth trajectory. Interesting relationships among the covariates and the initial starting levels, patterns of growth, and eighth grade reading outcomes were revealed, which will be discussed in greater depth below.

Lastly, three multiple group latent growth analyses were conducted. Original hypotheses for these analyses were not supported. First, results of the initial level analysis did not support the original hypothesis, such that results indicated no differences in reading achievement between the two groups at the start of kindergarten. Second, significant differences between the RD and LRA groups' patterns, or shapes of reading growth over time were found. However, exactly how the shapes differed could not be determined. The original hypothesis was that the two groups would have similar shapes but different rates of change. Given that the two groups had different shapes, no analyses were carried out on their rates of change. Third, it was hypothesized that the RD group would have significantly lower reading achievement scores at the end of eighth grade, but in fact, results revealed that the RD group had significantly higher reading achievement scores at the end of eighth grade.

Overall this study offers a unique contribution to the literature related to the differences in the longitudinal trends (i.e., kindergarten through eighth grade) in reading achievement between students with RD and LRA through the use of advanced statistical methods (i.e., LGM), control of important variables, consideration of special education eligibility, large sample size, and examination of initial reading levels at kindergarten entry, shape of reading growth, and eighth grade reading outcomes. A closer examination of the results follows.

Predicting Group Membership

The results of the logistic regressions indicated a number of important findings, some of which were in line with the current study's hypotheses, while others were not. First, as hypothesized, boys were more likely to be identified for the RD group than were girls. This is consistent with previous research that suggests that boys are more likely than girls to be referred for special education services (Anderson, 1997; Hibel, Farkas, & Morgan, 2010). Accordingly, in the current sample, given that both groups of children were struggling in reading, it is likely that other factors, such as externalizing behavioral problems, led boys in the RD group to be identified to receive special education services more often than girls. It is also possible that boys in the RD group had more severe reading problems than boys in the LRA group, especially at younger ages, given that girls have been found to develop reading skills more quickly than boys (Scheiber, Reynolds, Hajovsky, & Kaufman, 2015; Wolf & Gow, 1986).

Second, regarding both SES and race/ethnicity, the results indicated that children with lower SES were more likely to be identified for the LRA group, and racially/ethnically diverse students (i.e., African American, Hispanic, Asian, Other), as compared to White students, were more likely to be identified for the LRA group, which did not support this study's hypotheses. These results are also in line with the sample descriptive statistics that demonstrated that the LRA group was more racially/ethnically diverse with a lower SES than the RD group.

In support of these findings, researchers have found that racially/ethnically diverse and economically disadvantaged groups are in fact under-represented in special education, rather than over-represented, which is generally thought to be the case. Previous findings of over-representation of racially/ethnically diverse students in special education may have resulted from methodological limitations, specifically in regard to sampling procedures. That is, when over-representation is found, schools included in the study often have majority White student populations, so identifying a small number of ethnically/racially diverse students for special education services makes it appear like over-representation relative to the school's population.

In accordance with racially/ethnically diverse under-representation in special education, Hibel et al. (2010) demonstrated that within the ECLS-K sample, racially/ethnically diverse students were equally or less likely than White students to be placed into special education, and this finding held true when schools had a higher proportion of racially/ethnically diverse students. That is, schools with high enrollments of racially/ethnically diverse students were less likely to place students into special education. According to Hibel et al. (2010), one possible reason that racially/ethnically diverse students are less likely to be identified for special education is because they may more frequently attend disadvantaged schools with limited resources. Likewise, students who attend more disadvantaged schools may be less likely to stand out amongst similarly low-performing peers, which has been termed the frog-pond effect (Hibel et al., 2010). Therefore, in contrast to a significant amount of research that suggests that racially/ethnically diverse students are disproportionately over-represented in special education, the current study may be better aligned with research that suggests that racially/ethnically diverse students may be equally- or under-represented in special education at the national level. Please note, however, that although racially/ethnically diverse children were more likely to be in the LRA group in this study, not all of the students in this group were from racially/ethnically diverse backgrounds. Therefore, the current study did not have specific data to support the idea of the frog-pond effect, and the results of this logistic regression should not be over-generalized.

With further respect to SES, Hibel et al. (2010) also found that SES did not significantly predict special education identification, which contrasts with previous research that has shown that those with lower SES are more likely to be placed into special education (Blair & Scott, 2002). In addition, Siegel and Himmel (1998) found that children with higher SES were more likely to be diagnosed with RD, as a result of the relationship between IQ and SES, not as a result of a relationship between reading ability and SES. The authors suggest that children from lower SES backgrounds are less likely to be exposed to the same enriched environments that leads to higher IQ scores for those from higher SES backgrounds, and, therefore, those with RD from higher SES backgrounds may exhibit a greater discrepancy between their IQ and reading scores, leading to a diagnosis of RD and special education placement (Siegel & Himmel, 1998). In the current sample, because it is likely that a discrepancy model was used to identify students to receive special education services, those with higher SES may be included in the RD group due to a greater discrepancy between IQ and achievement. It is also possible that higher SES was predictive of RD group membership as a consequence of more advantaged and educated parents who were more involved in their child's schooling and advocated for a special education evaluation when they noticed that their child was struggling in reading (Barnard, 2004; Englund, Luckner, Whaley, & Egeland, 2004; Mann, McCartney, & Park, 2007) as well as that economically disadvantaged parents may have had less resources to support their child's learning and reading development outside of school (Davis-Kean, 2005). Furthermore, based on the results of Hibel et al.'s (2010) study, children attending disadvantaged schools with limited resources may not receive special education services due to a combination of both limited resources and the frog pond effect. Hence, in the present sample, children with lower SES were more likely to not receive special education services perhaps as a result of less discrepant IQ and achievement scores as well as limited parental and school resources; however, the current study did not have specific data to support this hypothesis, so caution must be taken when thinking about these findings.

Third, contrary to this study's hypothesis, the results of the logistic regression revealed that older children at kindergarten entry were more likely to be identified for the RD group than for the LRA group. This finding conflicts with Siegel and Himmel's (1998) study that found that children diagnosed with RD were younger than those who were identified as having LRA. One possible explanation for the conflicting findings is that the Siegel and Himmel (1998) did not specifically look at age at kindergarten entry, whereas the current study relied only on children's age at the fall of kindergarten. Another possible explanation for the conflicting findings is that children who entered kindergarten at older ages displayed more severe early learning/developmental concerns, and therefore, were more likely to be identified for special education services later in their schooling. It should be noted, however, that the RD and LRA groups' mean ages at kindergarten entry only differed in age by approximately two months, so although older children were statistically more likely to be identified for the RD group, perhaps there is not much clinical significance to this finding.

Growth Trajectory of the Full Sample

An unconditional latent basis model was the most appropriate and best fitting model for the data of the full sample. According to Bast and Reitsma (1998), a linear model of reading development does not take the reciprocal relationships between reading growth and other developmental processes into account (e.g., cognitive skills, behavioral changes), which also do not grow linearly. A linear model also assumes that the same reading processes that are important for early reading development (e.g., phonological awareness) remain important over time (Bast & Reistma, 1998). In other words, a linear model of reading development assumes that initial differences in reading achievement and subsequent reading growth do not change with development over time. Bast and Reistma (1997, 1998) also point out that growth models that incorporate the autoregressions of variables are useful in examining reading development over time. The current study did not use an autoregessive model, but the final unconditional model that achieved the best fit included the residual covariances/regressions of the reading achievement outcome variables, which are, in fact, autoregressive disturbances, or residuals. The autoregressive residuals, referred to as residual covariances/regressions throughout this manuscript, indicate that the observations at Time 2 depend on the observations at Time 1, and the observations at Time 3 depend on the observations at Time 2, and so on (Bast & Reistma, 1998). Using autoregressions does not mean that the measurements at each time point were unreliable (i.e., IRT scale scores); rather, there were influences that systematically changed one's reading skills between each time point that were not captured within the model itself (Bast & Reistma, 1998). Thus, the final unconditional latent basis model with residual covariances/regressions (i.e., autoregressive residuals) in the current study has been supported in previous literature.

Given that a latent basis model was used, it is important to discuss the basis coefficients and the pattern of reading growth over time. This basis coefficients of this model indicated that only 8% of reading growth occurred between the fall of kindergarten and the spring of first grade. The majority of growth (51%) in reading achievement occurred between the spring of first grade and the spring of third grade. Less growth, although still significant growth, occurred between the spring of third grade and the spring of fifth grade (28%), and even less growth, but also still significant occurred between the spring of fifth grade and the spring of eighth grade (13%). Accordingly, the greatest growth in reading achievement for this sample of poor readers occurred between the spring of first grade and the spring of third grade followed by slower growth from third to eighth grade.

These findings are relatively consistent with other research that suggests that the greatest growth in reading achievement occurs during the early elementary school grades and declines in later grades (Bloom, Hill, Black, & Lipsey, 2008; Cameron, Grimm, Steele, Castro-Schilo, & Grissmer., 2015; Hill, Bloom, Black, & Lipsey, 2007; Lee, 2010; Scammacca, Fall, & Roberts, 2015; Seltzer, Frank, & Bryk, 1994; Skibbe et al., 2008). For instance, Bloom et al. (2008) found that the mean effect size across seven standardized reading tests decreased from 1.52 for the difference between the spring of kindergarten and the spring of first grade, to 0.36 for the spring of third grade to the spring of fourth grade, and then to 0.19 for the spring of ninth grade to the spring of tenth grade. This decelerating growth pattern of reading achievement has been found across studies that examine multiple areas of reading development (e.g., decoding and comprehension; Bloom et al., 2008; Hill et al., 2007; Lee, 2010; Scammacca, Fall, & Roberts., 2015) as well as studies that focus on only one domain of reading development (e.g., comprehension; Seltzer et al., 1994). Furthermore, some of these studies utilized latent growth modeling to examine the reading growth trajectories (Cameron et al., 2015; Skibbe et al., 2008), while other studies measured growth by looking at effect sizes (Bloom et al., 2008; Hill et al., 2007; Scammacca et al., 2015). Across the different methodologies of these studies, results support a pattern of accelerated reading growth in early grades and decelerating reading growth in later grades.

Some researchers attribute this pattern of growth to a reduced capacity for acquiring new skills and knowledge as children age as well as the increased difficulty of academic demands and expectations in later grades (Lee, 2010). It may also be possible that as children age and enter middle school, reading instruction, including comprehension instructions becomes less of a focus, because students are expected to have the skills to read and comprehend the material. Consequently, in combination, reduced capacities for acquiring new knowledge, increasing school demands, and changing instruction support an increasing pattern of reading growth in early school grades followed by deceleration in growth as children get older.

However, it is also important to point out that this study did not measure the same reading skills over time, such that the early assessment time points were evaluated using phonological awareness and decoding skills, whereas the later assessment time points were based on more advanced reading skills, such as comprehension and critical thinking. Therefore, the deceleration observed in the current study may reflect a lack of skill growth specifically related to reading comprehension and higher-order reading skills (e.g., critical analysis), not just a lack of progress in early reading skills, such as phonological awareness and decoding. Although reading instruction becomes less of a focus, expectations increase, and new knowledge may be harder to acquire in later school grades for all children, some children with reading problems are likely provided with accommodations (e.g., read aloud, speech to text) in the classroom that allowed them to still access more advanced content despite poor decoding and continue to progress in their reading comprehension and critical thinking skills. That is, when accommodations are provided to children struggling with learning to read, some of the deceleration that tends to occur in the general population of students may be prevented.

Although the current study found generally similar results to previous studies regarding the pattern of reading growth (i.e., increasing pattern in early school grades, followed by deceleration in later grades), there is an important nuance to point out. In the present study, the majority of reading growth (51%) occurred between the spring of first grade and the spring of third grade, while minimal growth occurred during kindergarten (8%). However, previous research has found that a significant amount of growth occurs before and during kindergarten, and growth slows from thereafter (Hill et al., 2007; Skibbe et al., 2008). These studies that have found significant growth before and during kindergarten did not specifically examine the reading growth of children with RD and LRA reading achievement. Hill et al. (2007) studied the reading growth in national norming samples of six standardized reading tests, which likely included children with RD and LRA, but they were not the majority or focus of the sample. Likewise, Skibbe et al. (2008) examined the reading growth trajectories of children with language delays, and although language impairments are highly comorbid with RD, not all children with language impairments have reading problems. Thus, it is possible that the reading growth trajectories of children with RD and LRA are slightly different than those of the general population or those with language delays.

Accordingly, Scammacca et al. (2015) investigated reading growth using national norming samples and were specifically interested in the reading growth (i.e., effect sizes) of those whose scores fell below the 25^{th} and 10^{th} percentiles. Children who scored below the 25^{th}

and 10th percentiles had an average effect sizes of 1.23 and 1.09 for reading growth from kindergarten to first grade, respectively, whereas those who scored above the 50th percentile had an average effect size of 1.42 from kindergarten to first grade (Scammacca et al., 2015). As such, it appears that those with scores in the lower performing groups made less growth from kindergarten to first grade than those in the typically performing group, which may support the findings of the current study where students with reading problems did not make as much growth from the fall of kindergarten to the end of first grade as other researchers have reported.

Furthermore, Scammacca et al. (2015) found that for those who scored above the 50th percentile, a clear pattern of deceleration after kindergarten emerged, such that from kindergarten to first grade the effect size was 1.03, which ultimately dropped to 0.21 from seventh to eighth grade and to 0.11 from tenth to eleventh grade. For those below the 25th and 10th percentiles, a slightly different pattern emerged. In contrast to a steep pattern of deceleration, the reading growth of those who scored below the 25th and 10th percentiles did not decelerate as quickly (Scammacca et al., 2015). For example, between first and second grade, those who scored below the 10th percentile made just as much progress as they did from kindergarten to first grade (i.e., first to second grade: d = 1.10; kindergarten to first grade: d = 1.09), and those who scored below the 25th percentile did not seem to decelerate as much from first to second grade as did those who scored above the 50th percentile (i.e., 25th percentile: d = 1.10; 50th percentile: d = 1.03; Scammacca et al., 2015). In addition, the effect sizes for children in the 25th and 10th percentile groups (i.e., 25th percentile: d = 0.66; 10th percentile: d = 0.74) were greater than the effect sizes for the children in the 50th percentile group from second to third grade (i.e., d = 0.56; Scammacca et al., 2015). Whether these differences in effect sizes were statistically significant is unknown. Nonetheless, it is possible that the current sample of children with RD and LRA have slightly different reading growth patterns than average readers, such that they experience the greatest growth in reading skills between the end of first grade and the end of third grade, followed by a deceleration in growth from the end of third grade through the end of eighth grade.

The results of the current study are also relatively consistent with those found by Cameron et al. (2015) who studied the reading growth trajectories using advanced non-linear, S-shaped growth models (e.g., Gompertz curve) in two different national datasets (i.e., ECLS-K, Longitudinal Survey of Youth – Children and Young Adults [NLSY-CYA]). Using non-linear models such as the Gompertz curve to model reading achievement data is based on the assumption that reading growth "is expected to follow an S-shaped curve, with an initial gradual effortful process of basic skill acquisition, followed by rapid improvement in fluency and comprehension as phonics and sight word recognition become automatized. After this point, less rapid improvement, or a leveling-off phase ensues, where a host of other factors beyond decoding begin to contribute to children's expanding literacy repertoires" (Cameron et al., 2015, p. 791). An S-shaped model allows for the examination of total change, rate of change, and timing of change, or when children experience the fastest growth and greatest deceleration in growth. Results of Cameron et al.'s (2015) study indicated that in the ECLS-K dataset, children experienced 37% of their total growth by the middle of first grade and were experiencing the greatest growth at this time. In the NLSY-CYA dataset, children experienced 37% of their total growth by the middle of second grade and were experiencing the greatest growth at this time (Cameron et al., 2015).

Visual analysis of the predicted unconditional latent basis model (see Figure 4) seem to suggest a naturally-occurring S-shaped model, in which students experienced the greatest growth in reading skills between the end of first grade and the end of third grade. These results, therefore, are relatively consistent with those of Cameron et al. (2015), and suggest that as children are "learning to read" (i.e., decoding, fluency, and comprehension), they experience the greatest growth in their reading skills, but when the focus shifts to "reading to learn" (Chall, 1983), growth in reading skills slows down.

Still, it is important to consider why children with RD and LRA only experienced 8% of growth in reading achievement from the fall of kindergarten to the spring of first grade, when other studies have found the greatest growth in reading before (e.g., 54 months of age) and during kindergarten (Hill et al., 2007; Skibbe et al., 2008). First, as described above, both of these studies did not specifically study students with RD and LRA. Second, Hill et al. (2007) did not utilize latent growth modeling; rather, they only developed effect sizes from standardized reading tests. As such, IRT scale scores, which are the most appropriate way to measure change over time (Seltzer et al., 1994), were not utilized. Third, differences in the time periods studied both in the current study and in Hill et al. (2007) and Skibbe et al.'s (2008) studies may have influenced the discrepant findings. In the current study, children's reading achievement prior to kindergarten grade was not examined. However, Skibbe et al. (2008) studied reading achievement beginning at 54 months of age, and then measured reading achievement again in the first, third, and fifth grades. Perhaps, the time between 54 months of age and kindergarten is a crucial time period to consider in reading development, and had this time period been included in the current study, maybe the growth pattern would have looked different in the early school years. Similarly, Skibbe et al. (2008) did not examine reading growth past fifth grade, and because latent growth models fix the first and last time points to zero and one, respectively, it is possible that reading growth would have been distributed differently if their time frame was extended. Finally, Hill et al. (2007) and Skibbe et al. (2008) did not explicitly consider the ways in which individual differences and other important variables (e.g., gender, SES, race/ethnicity, age) influence reading growth trajectories over time.

Lastly, the graphical representations of individual differences in the unconditional model (i.e., spaghetti plots) revealed great individual variation in the reading achievement scores by the eighth-grade time point (see Figure 5). Previous research indicates that individual differences in reading achievement increase as children get older (Bast & Reistma, 1997, 1998; WIliamson, Applebaum, & Epanchin, 1991). This is supported in the current study, such that there was relatively less individual variation in reading achievement, although still significant, at the start of kindergarten, and this individual variation increased, or spread out over time, which is called a fan-spread effect (Bast & Reitsma, 1997, 1998; Aarnouste & Van Leeuwe, 2000). In support of this fan-spread effect, previous research has found evidence of a Matthew effect (Bast & Reistma, 1998; McKinney & Feagan, 1984; McNamara et al., 2011), which is a model originally proposed by Stanovich (1986) to explain individual differences in reading development over time. Based on the Matthew effect model, while individual variation increases over time, rank ordering of individuals are relatively stable, such that those with the poorest initial reading achievement make the least gains and maintain their positon at the bottom of the fan, whereas those with the highest initial reading achievement make the most gains and represent the higher positons in the fan (Bast & Reistma, 1998). Although there appears to be evidence of a fan-spread effect in the current sample, there does not appear to be true evidence of a Matthew effect, specifically regarding rank order stability. In other words, because the correlation and covariance between the initial level and slope were not statistically significant, there is neither evidence that those with the highest initial reading achievement maintained their position and ended up with the highest eighth-grade reading achievement nor that those with the lowest initial reading achievement maintained their position and ended up with the lowest eighth-grade reading achievement. A significant and positive correlation and covariance between the initial level and slope would have supported a Matthew effect in the current study.

Similarly, as a result of the non-significant initial level and slope correlation and covariance, the results of the unconditional model for the full sample also do not support the compensatory trajectory theory of reading development, which is the idea that children with initial lower reading achievement progress at a fast rate than those with higher initial reading achievement (Aarnoutse et al., 2001; Aunola et al., 2002; Parrila et al., 2005; Phillips et al., 2002; Scarborough, 2003). A significant and negative correlation and covariance between the initial level and slope would have supported the compensatory trajectory theory of reading development. Accordingly, within this sample of poor readers, which included both identified children with RD and non-identified children with LRA, there is limited evidence to suggest that a specific relationship exists between initial levels of reading achievement and growth over time.

Nonetheless, it is important to point out that a group of typically developing, or average readers, was not included in the current sample. Given that the current sample comprises only poor readers, drawing conclusions about a Matthew effect and a compensatory trajectory is not that meaningful. Consideration of a Matthew effect and compensatory trajectory in the current sample is asking whether those with the poorest of the poor reading skills fall further behind than or catch up to their peers who also have poor reading skills, which may seem like a rather circuitous exploration. It should also be noted that some researchers argue that the analysis of Matthew effects and the interpretability of results are generally questionable due to psychometric and scaling issues related to the measurement of achievement (Protopapas et al., 2016). Still, attempting to answer such questions related to a Matthew effect and the compensatory trajectory would have been more appropriate had a group of average, non-impaired readers been included in the sample. Nevertheless, comparisons of the patterns of growth between the RD and LRA groups are still necessary to consider and will be further discussed below. Before discussing the multiple group LGMs, however, other variables aside from reading achievement itself that may have accounted for the individual differences in reading achievement at the start of kindergarten, over time, and at the end of eighth grade must be examined.

Covariates and Growth Trajectory in the Full Sample

To examine the influence of important demographic variables on initial reading achievement, growth over time, and eighth grade outcomes, gender, race/ethnicity, SES, and age were included as covariates in the unweighted conditional model. The sections below will discuss the results and the significance of each of the covariates.

Gender

In the current sample of children with RD and LRA, the results of the unweighted conditional model indicated that girls had significantly higher reading achievement in the fall of kindergarten as well as at the end of eighth grade. However, there were no differences between boys and girls in terms of total growth, or the shape of reading growth, from kindergarten through eighth grade. In other words, girls had an initial advantage in reading development at the start of kindergarten and maintained this initial advantage over time. Boys in the sample experienced similar reading growth compared with girls, but did not experience enough growth to help them to catch up to the girls by the spring of eighth grade.

The finding that girls have initially higher reading achievement in the fall of kindergarten is consistent with research that has shown that girls develop reading skills more quickly than boys (Morgan et al., 2008b; Scheiber et al., 2015; Wolf & Gow, 1986). Furthermore, there is evidence to suggest that girls continue to maintain their advantage in reading in fourth and eighth grades (NCES, 2016; Scheiber et al., 2015). It should be noted that some researchers propose that girls may be more likely than boys to experience late-emerging RD and/or reading comprehension deficits (Adolf et al., 2010). If these girls experienced reading problems after fifth grade (i.e., either RD identification or LRA), then they were not included in the current sample. Consequently, while the findings related to gender and reading achievement in the current study are generally consistent with previous research, one must consider the way in which the sample was defined and reading achievement was measured in the current study.

Race/Ethnicity

Results of the unweighted conditional model revealed several important findings regarding the relationship between race/ethnicity and reading achievement. First, Hispanic, Asian, and other racially/ethnically diverse children were found to have significantly lower reading achievement scores at the fall of kindergarten compared to White children, whereas African American children's reading achievement was not found to significantly differ from White children in the fall of kindergarten. Other researchers who examined achievement growth trajectories within the entire ECLS-K sample found similar results, except that Asian children had significantly higher initial reading scores (McCoach et al., 2006; Morgan et al., 2008b). Second, African American and other racially/ethnically diverse children had significantly less total growth in their reading achievement compared to White children; however, Hispanic and Asian children did not significantly differ from White children in total growth. These results are partially supported by Morgan et al.'s (2008) study, which found that all of the racially/ethnically diverse groups had slower reading achievement growth compared to White children. Third, in the spring of eighth grade, African American, Hispanic, and other racially/ethnically diverse children performed significantly lower than White children on the reading assessment, and there were no differences between the reading achievement scores of White children and Asian children in the spring of eighth grade.

Generally, there is evidence to suggest that African American, Hispanic, and American Indian/Alaska Native children have significantly lower academic achievement than White children throughout elementary and middle school, whereas Asian children tend to have higher academic achievement than White children throughout elementary and middle school (McCoach et al., 2006; Morgan et al., 2008b; NCES, 2016). Notably, in the current study, and also consistent with McCoach et al. (2006) and Morgan et al.'s (2008b) study, with SES controlled, African American children and White children did not differ in their reading skills in the fall of kindergarten. Thus, why African American and White children did not differ in the fall of kindergarten may have resulted from similar effects of early reading experiences (e.g., preschool, home literacy environment) on their kindergarten reading skills that did not become differentiated until receiving formal reading instruction in elementary school. It may also be that SES accounts for much of the variance in the differences between the reading skills of African American and White children in the fall of kindergarten, such that once SES is controlled, any significant differences between these two groups disappear. Further discussion about the relationship between SES and reading achievement follows in a below section.

In the current sample, both Hispanic and Asian students, however, had significantly lower reading achievement in the fall of kindergarten compared to White students. Lower reading achievement in these two groups, especially considering they are poor readers, may have resulted from being bilingual, being an English language learner (ELL), or having limited English proficiency. Research suggests that children who are learning two languages or are bilingual tend to have slower language development than monolingual children (Hoff & Core, 2015). Accordingly, language-related factors may account for a significant amount of the variation in Hispanic and Asian children's reading skills in the fall of kindergarten, especially because neither the spoken language nor language ability were included in the current study as covariates.

Even though Asian and Hispanic children had lower initial reading skill levels, they demonstrated similar reading growth compared to White children over time. For Hispanic children, their similar growth, albeit lower eighth-grade reading outcomes, may have resulted from a complex relationship between ELL-status and reading instruction. Hispanic children may be less likely to receive reading instruction prior to kindergarten, such as in preschool settings (Ansari, 2017), so once they entered formalized schooling, they may have experienced significant language and reading growth. However, because many Hispanic students may have attended disadvantaged and segregated schools with limited resources (Hibel et al., 2010), though they had similar growth compared to White students, they never actually caught up to them by the end of eighth grade.

Asian children not only demonstrated similar growth but also had similar eighth-grade outcomes compared to White children, which may be explained in a few ways. First, research suggests that as a group, Asian children tend to have higher academic performance than White children, and there tends to be high value placed on education and achievement within many Asian cultures (e.g. Goyette & Xie, 1999). Second, combined with the high value placed on achievement, Asian parents may be more likely to attempt interventions aimed at improving their child's reading skills, which may result in increased reading growth (Hibel et al., 2010). Third, Asian children in the current sample may attend schools within relatively higher levels of academic achievement, more resources, and an overall higher SES, which may foster greater reading growth. Likely, it is a combination of these explanations that led to greater reading growth and better eighth grade reading outcomes for Asian children in the current study.

In contrast to Asian and Hispanic students, both African American and other racially/ethnically diverse children made less progress in reading development over time and had lower eighth grade reading outcomes compared to White children. Although individual SES was controlled in the current study, school SES was not controlled. Taking a bioecological perspective (Bronfenbrenner & Ceci, 1994; Bronfenbrenner & Morris, 1998; Bronfenbrenner & Morris, 2006), one must consider not only the child's cognitive and reading skills but also environmental factors from multiple levels that may explain such findings. At broader, macrosystemic and exosystemic levels, one must consider factors such as institutional racism and policies that discriminate against those from racially/ethnically diverse groups (e.g., segregated housing, segregated schools, legal system). Racism and discriminatory practices also extend down from the broader macrosystem to the mesosystem and microsystem. For example, at the mesosystemic level, a child's school who does not use culturally sensitive practices and materials to communicate with parents may prevent the formation of strong home-school relationships, which in turn can influence academic achievement as well as high school dropout and completion rates (Barnard, 2004). At the microsystemic level, teacher and instructional quality tend to be weaker in disadvantaged and segregated schools, and some teachers may hold lower expectations or perceptions of academic potential for racially/ethnically diverse children, which also influences academic achievement (Kreisman, 2012). When racially/ethnically diverse children are faced with discrimination, racism, and less effective instructional and classroom practices that directly influence their school performance, one can see why African American and other racially/ethnically diverse children may make less academic growth over time and never catch up to their White peers.

Socioeconomic Status (SES)

Not surprisingly, results of the unweighted conditional model showed that children with lower SES had significantly lower initial reading achievement scores, made significantly less gains over time, and had significantly lower reading achievement scores at the end of eighth grade. These findings are supported by Morgan et al. (2008b) and McCoach et al. (2006) who also found that SES was positively related to initial reading status and rate of growth over time. Generally, research supports the relationship between SES and reading achievement, such that both income and parental education are important predictors of children's reading skills (Smith, Brooks-Gunn, & Klebanov, 1997) and early measures of SES predict children's later reading scores throughout their education (Jimerson, Egeland, & Teo, 1999; Molfese et al., 2003). Furthermore, children from economically disadvantaged homes acquire oral language skills more slowly, exhibit delayed letter recognition and phonological sensitivity, and are at risk for reading difficulties (Hoff, 2013; Whitehurst & Lonigan, 1998), and reading outcomes have been found to be significantly related to certain factors of family context, such as home-literacy environment, number of books owned, parental distress, and receipt of center-based care (Aikens & Barbarin, 2008). Accordingly, in the current study children with from lower SES backgrounds had reduced initial reading scores and made less growth in reading skills over time.

In addition to individual SES factors, school and classroom level SES-related factors may also influence students' reading achievement. One study found that students' classrooms (e.g., race/ethnic composition, free/reduced-lunch eligibility, teacher experience) contributed to approximately 60% of the reliable variance in students' reading growth in a given year (Rowan, Correnti, & Miller, 2002). Thus, school and classroom related variables may play a crucial role in students' reading achievement. More specifically, school SES, which is usually measured by the percentage of students who are eligible for free/reduced-lunch prices at a school (Sirin, 2005), influences students' reading achievement. Students who attend schools with lower school SES are more likely to face barriers and challenges that affect their learning environments at both the societal and school levels. That is, coming from a bioecological perspective (Bronfenbrenner & Ceci, 1994; Bronfenbrenner & Morris, 1998; Bronfenbrenner & Morris, 2006), it is crucial to consider environmental and contextual factors that are related to SES, such as school culture, classroom management, instructional practices, and school resources, that influence students' academic achievement. For example, Wenglinesky (1998) found that lower-SES schools had poorer instructional practices, limited materials, teachers with less experience, and higher student-teacher ratios. Other researchers suggest that modifiable school characteristics (e.g., resources, instructional practices), peer effects, and schools' responses to the student composition (e.g., lowering expectations for students from low SES backgrounds) are the three school-related mechanisms that influence student achievement (Rumberger & Palardy, 2005). Even broader, the location of the school has been found to influence the amount of resources available per students, and such available resources then in turn influence students' academic achievement (Unnever et al., 2000). Consequently, the effects of SES on students' reading achievement are important to consider at not only the individual level but also the school and neighborhood levels, as both may influence students' initial levels of reading achievement and progression of reading development over time.

Age

In the present study, results of the unweighted conditional model indicated that older children had significantly higher scores in the fall of kindergarten, but then older children made significantly less progress in their reading achievement over time. By the end of eighth grade, age did not significantly predict reading achievement scores. These findings are consistent with other research that suggests that older children enter kindergarten with higher reading scores, yet their advantage tends to disappear by the end of elementary school (McCoach et al., 2006; Stipek & Byler, 2001). Wu, Morgan & Farkas (2014) also found that older children had initially higher reading achievement over kindergarten, but they had slower growth over time. Similarly, in the current study, older children not only lost their advantage over time but also made significantly less growth in reading skill over time. It is not entirely clear as to why the older children made less progress than younger children, but possible explanations include differences at the school and instructional levels as well as that those who entered kindergarten at later ages had more impairing reading problems, developmental delays, and/or lower IQs. Regardless of why older children made less growth over time, there were no significant differences between older and younger students' reading achievement by the end of the eighth grade, as suggested by other research (Stipek & Byler, 2001).

Group Comparison of Initial Level of Reading Achievement

The first important finding from the unweighted multiple group LGM is that the RD and LRA groups did not differ in terms of their initial reading achievement scores at the fall of kindergarten. In other words, the two groups, both defined by reading problems, had similar reading skills at the start of kindergarten before being identified as either in need of special education services or having poor reading achievement. Based on the literature review, this was the first study to utilize the fall of kindergarten as the initial measurement, or intercept. Nonetheless, the two longitudinal studies that examined differences in the reading achievement between children with RD and children with LRA using slightly later initial levels also found no differences between the two groups reading achievement. Francis et al. (1996) used age eight as the initial level, and O'Malley et al. (2002) used the beginning of first grade as the initial level. It is difficult to tell whether the two groups in the current study would still have had similar levels of initial reading achievement if a different intercept (e.g., age eight, first grade) had been used. Looking at Figure 6, it appears that the two groups diverged slightly in their reading achievement by the spring of first grade, and by third grade had dissimilar patterns of reading achievement. Still, Francis et al. (1996) and O'Malley et al. (2002) did not define their groups using special education eligibility, which in the current study, may have contributed to the diverging patterns before third grade.

Nonetheless, it is not surprising that the two groups appeared similar in terms of reading achievement at the beginning of kindergarten. First, Shaywitz et al. (1992a) demonstrated that reading achievement exists on a normal curve, or a continuum, such that those with RD are not a distinct group; rather, reading problems vary in severity, and those with the greatest reading difficulties fall on one end of the continuum. Thus, in the present sample of children who all displayed some form of reading problems between first and fifth grade, it makes sense that the two groups would have had similar levels of reading achievement right before receiving formalized reading instruction or intervention in elementary school, especially with gender, race/ethnicity, SES, and age controlled. Moreover, although students in the RD group may have gone on to develop significant discrepancies between their achievement and IQ or did not respond to intervention, there is nothing that would suggest that the RD and LRA groups would have had qualitatively different early reading achievement, based on the research that suggests that deficits in phonological awareness are the foundation for most early reading problems (Stanovich, 1985, 1988; Stuebing et al., 2002).

Further, research suggests that differences in academic achievement in kindergarten may be related to differences in children's early literacy experiences. Previous research suggests that children with higher quality preschool experiences and more enriching home-literacy environments have higher achievement in early elementary school grades (Pianta & McCoy, 1997; Snow et al., 1998); however, it possible that given the lower SES and racially/ethnically diverse sample in the current study, children in both the RD and LRA groups had relatively similar early literacy experiences, which in turn affected their reading development by the fall of kindergarten in similar ways. It is also possible that because SES was controlled in the multiple group LGM, individual and group differences in reading achievement that would have resulted from differences in SES and early literacy experiences (e.g., preschool) were accounted for within the model. It should also be noted that in a sample that was similar to the current sample in terms of racial/ethnic composition, in which SES, race/ethnicity, gender, age, and other variables were controlled, Wu et al. (2014) found that a group of students with the lowest reading and math achievement from kindergarten through fifth grade did not significantly differ from a group students with typical reading and math growth at the start of kindergarten in terms of reading achievement, although the low achievement group fell behind over time. Accordingly, it may not only be that children with RD and LRA display similar levels of reading achievement in the fall of kindergarten but also that children in general tend to display similar levels of reading achievement in kindergarten when important variables are considered (e.g., SES, race/ethnicity, gender, age, approaches to learning, school type).

Sensitivity Analyses

Notably, the results of the sensitivity analyses for the initial level group comparison were similar to the results using the original groups. That is, defining the LRA group using a third-grade only reading achievement score below the 25th percentile, a RD group that included those identified for special education services in kindergarten, a RD group that excluded children identified for special education services in fifth grade, and a RD group that excluded children identified for special education services in fourth and fifth grade did not alter the results. Accordingly, regardless of how the groups were defined, the results consistently indicated no differences between the RD and LRA groups' reading achievement at the beginning of kindergarten, which is also fitting with the results of Wu et al.'s (2014) who did not find differences between typical readers and children with reading problems in the fall of kindergarten.

Group Comparison of Shape, or Pattern of Reading Achievement

The second important finding from the unweighted multiple group LGM revealed significant differences between the two groups' patterns of reading achievement growth over time. The two groups had relatively similar growth from the fall of kindergarten to the spring of first grade (i.e., approximately 10% of growth). However, the patterns of reading achievement diverged from the spring of third grade onwards. The RD group experienced 27% of growth between the spring of first grade and the spring of third grade, 19% of growth between the spring of third grade and the spring of fifth grade, and 42% of growth between the spring of fifth grade and the spring of eighth grade. On the other hand, the LRA group experienced growth greater than 100% between the spring of first grade and the spring of fifth grade, before going back down at the final eighth-grade time point. For the LRA group, growth from the spring of third grade to the spring of fifth grade was not as steep as the growth from the spring of first grade to the spring of third grade. Importantly, while the LRA group experienced deceleration in their reading achievement between fifth and eighth grade, the RD group experienced acceleration in their reading development from fifth to eighth grade. In sum, the RD group experienced acceleration in their reading scores until the spring of third grade, then a slight deceleration from the spring of third grade to the spring of fifth, and then a more rapid acceleration from the spring of fifth grade to the spring of eighth grade; the LRA group experienced gradual acceleration in their reading achievement until the spring of first grade, and then steep acceleration from the spring of first grade to the spring of fifth grade, and then deceleration from the spring of fifth grade to the spring of eighth grade.

Previous research that examined the longitudinal differences in the reading achievement between those with RD and those with LRA did not find group differences in the slopes, or rates of change over time (Francis et al., 1996; O'Malley et al., 2002). These studies, however, did not utilize a latent basis model to examine the patterns of reading development over time, and these studies did not consider whether students in their samples were receiving special education services. Had Francis et al. (1996) and O'Malley et al. (2002) utilized a similar statistical modeling approach and taken special eligibility identification into consideration, their results may have turned out differently. Thus, this study extends the literature by looking more deeply at patterns of reading growth over time using a non-linear model, not just the rate of change over time in a linear or quadratic model.

Consideration of Reading Processes and Sub-Processes

As a result of the groups' differing patterns, one must consider potential reasons for why the differences exist. One possible explanation is that because different reading-related skills were assessed at each time point, the groups experienced growth in different reading-related processes at different points in time. Although IRT scale scores allow for the comparison of change over time even when the items on the test change at each measurement point, IRT scale scores do not permit for the analysis of which specific reading processes (e.g., decoding vs. fluency vs. comprehension) and sub-processes (e.g., phonological awareness vs. listening comprehension) were developing over time. Consequently, given the way the groups were defined in the current study, both the RD and LRA groups included students with different types of reading deficits, not only deficits in phonological awareness. Therefore, the combination of diverse reading deficits, the reading processes measured, and the developmental time points at which measurement occurred may have led to differential patterns in reading achievement between the two groups.

In support of the idea regarding the heterogeneity within each group as well as differences in how reading processes developed depending on the type of reading deficit, researchers have identified that certain reading processes are more predictive of certain reading problems than other reading processes. For instance, previous research suggests that children with reading comprehension deficits do not necessarily have deficits in phonological awareness; rather, they have deficits related to semantics and higher-order language processes (Nation, Clarke, Marshall, & Durand, 2004). Furthermore, Adolf et al. (2010) identified different variables that best predict reading impairments in second grade versus eighth grade. In second grade, letter identification, sentence imitation, and mother's level of education or rapid naming were the best predictors of whether a child had a reading impairment, and in eighth grade, phoneme deletion, sentence imitation, nonverbal IQ, mother's education level, grammatical completion, and rapid naming were the best predictors of whether a child had a reading impairment (Adolf et al., 2010). Accordingly, in eighth grade, more predictors were needed to model whether a child would have a reading impairment, and two of these additional predictors (i.e., nonverbal intelligence, grammatical completion) were unique to the eighth-grade model. These authors suggest that the differences in the predictor variables at second and eighth grade mean that there are qualitatively different reading impairments in second grade versus eighth grade, such that those in identified second grade tend to have deficits related to decoding skills, whereas those identified in eighth grade tend to have related to broader language ability and higher-order thinking deficits (Adolf et al., 2010). Other researchers have also shown that poor comprehenders may have more language-based deficits than those who are considered poor decoders (Geva & Massey-Garrison, 2013; Tong, Deacon, Kirby, Cain, & Parrila, 2011). In further support of this idea of different types of reading impairments at different grade levels, Adolf et al. (2010) found that about 42% of children who were classified as reading impaired in eighth grade were not considered reading impaired in second grade. Similarly, approximately 49% second graders who were classified as reading impaired were considered good readers in eighth grade (Adolf et al., 2010).

Therefore, in the current study, the RD group may have experienced significant gains in reading achievement from fifth to eighth grade because their primary deficits were in phonological awareness, and by eighth grade, with accommodations (e.g., having the text read to them), they were able to comprehend and read text for meaning well. In other words, even if the two groups experienced growth in different reading-related processes at different time points and consisted of students with different types of reading impairments, one must also consider that those in the RD group were likely given specific accommodations (e.g., read aloud, text to speech) both in the classroom and on the reading assessment itself to help them overcome their reading difficulties (e.g., phonological awareness, decoding), continue to learn more advanced reading skills, such as comprehension, and access the content, as they continued on in their education. Thus, when measuring different reading skills at different grade levels, it makes sense that those identified in earlier grades with RD may no longer be considered poor readers, or at least be on a path of continued growth in later school grades, which is consistent with Adolf et al.'s (2010) findings.

In contrast, for the LRA group, which was more racially/ethnically diverse and from lower SES backgrounds as compared to the RD group, their primary deficits may have been in reading comprehension and broader language-based skills, as opposed to phonological awareness and decoding deficits. Hence, the LRA group may have benefited from formalized reading instruction from first through fifth grade, which allowed their decoding and fluency skills to progress, but by eighth grade when the direct reading assessment was primarily measuring reading comprehension skills, they did not perform as well as the RD group. These students in the LRA group were also likely not given classroom accommodations to help them access the content and continue to make progress. Further, children from disadvantaged families with limited resources are less likely to experience academic opportunities outside of school and verbally rich environments than are children from higher SES families (Davis-Kean, 2005). Generally, research supports the idea that the language skills of children from racially/ethnically diverse and from low SES backgrounds decline over time, but it is commonly believed that this deceleration begins around fourth grade when the switch between "learning to read" and "reading to learn" occurs (Chall & Jacobs, 2003; Chall, Jacobs, & Baldwin, 1990; Kieffer, 2010). However, these hypotheses as to why children in the LRA group experienced a decline in their reading skills in this study are only conjecture, as specific information about the classroom environments and instruction of students in this sample was not available.

Specifically for the RD group, the results of the current study are more in line with those of Etmanskie et al. (2016) who found that children who experience a “fourth grade slump” in reading development, or a decline in reading skill when the switch between “learning to read” and “reading to learn” occurs, still tend to improve in their reading skills after fourth grade. In the current study, this is consistent the RD group’s pattern of reading development, such that they experienced slightly less growth between the spring of third grade and the spring of fifth grade (i.e., 19% of growth) than they made between the spring of first grade and the spring of third grade (i.e., 27% of growth), but they continued to make significant progress from the spring of fifth grade through the spring of eighth grade (i.e., 42% of growth). The progress the RD group made in later grades may be in part due to the accommodations and support they received to help them continue to learn to comprehend material and access more advanced curriculum, despite their early reading challenges. The LRA group, on the other hand, may have experienced a “fourth grade slump,” but they also still experienced positive growth from the spring of third grade to the spring of fifth grade. By the end of fifth grade, the LRA group experienced significant deceleration in their reading growth. Thus, it is possible that children with RD continued to benefit from reading instruction/intervention and special education services, including accommodations, throughout elementary school and middle school, whereas those in the LRA group either may not have benefited as much from the instruction/intervention they received or may not have received enough support and accommodations to continue to progress their reading development through the end of elementary school and then middle school.

Consideration of Instructional and Environmental Factors

In combination with the complex relationship between reading-related skills and students' deficits in those skills over time, it is necessary to consider how students' schools and instructional environments from an ecological perspective may have contributed to differences in the two groups' reading achievement patterns over time. In the current sample, students in the RD group received special education services at some point between first and fifth grade. Although research tends to suggest that children who receive special education services for having RD or other disabilities do not typically catch up to their same-age peers in reading achievement (Morgan et al., 2010, 2011), the current study did not allow for the comparison between the RD group and a group of non-impaired readers. Nevertheless, while it is possible that the RD group in the current study may not have caught up to non-impaired readers, students in this group experienced growth in their reading development from kindergarten through eighth grade, with only slightly slower growth between third and fifth grade. It should also be noted that those in the RD group received instruction via not only their special education placements but also their general education.

Regarding the instructional practices and environment in general educations classrooms for both the RD group and the LRA group, the current study does not allow for any conclusions to be drawn about type of accommodations, instruction, interventions, and environmental factors that contributed to differences between the RD and LRA groups' patterns of reading achievement over time. Still, it is important to think about what type of reading practices may be most beneficial for students in both general education and special education settings. For example, Connor and colleagues (Connor, Morrison, & Katch, 2004a; Connor, Morrison, & Petrella, 2004b; Connor et al., 2009) demonstrated the importance of individualized and differentiated reading instruction. In general, they have found that whether children benefit from code-based instruction (i.e., instruction that emphasizes phonological awareness, phonemic awareness, and decoding) versus from comprehension-based instruction (i.e., instruction that focuses on reading activities and language development) depends on the child's initial strengths and weaknesses and instructional time devoted to such activities (Connor et al., 2004a, 2004b, 2009). For example, children with lower initial decoding skills may benefit more from code-based instruction, whereas those with adequate decoding skills, but low vocabulary benefit more from code-based instruction with increasing amounts of comprehension-based instruction as the year progresses (Connor et al., 2004a, 2004b, 2009). In order to provide students with the most effective reading instruction, teachers must assess and understand where their students' initial reading skills lie and then work to differentiate instruction within their classroom, so that the learning needs of all students are met. Overall, although the specific type of reading instruction the students in both the RD and LRA groups received is unknown, it is possible that the students in the RD group received more individualized instruction based on their needs as well as accommodations, which allowed them to have a steadier trajectory of growth. Further, individualized instruction and accommodations not only helped those in the RD group continue to improve in their reading skills but also helped them to continue to access increasingly more difficult content in spite of their reading challenges. On the other hand, the LRA group likely did not receive accommodations in earlier grades, which prevented them from developing stronger comprehension skills and accessing more advanced content, and they likely received less individualized instruction throughout their education, perhaps leading to a decline in their reading growth after fifth grade.

Moreover, in addition to individualized reading instruction, other environmental factors may have played a role in the different reading achievement patterns between the two groups. Those in the LRA group were from more racially/ethnically diverse and disadvantaged backgrounds compared to the RD group, so variables such as parent involvement and home-school relationships, school resources, neighborhood violence, housing issues, etc. may have influenced the reading achievement trajectories of the LRA group differently than for the RD group, even though related variables, such as individual SES and race/ethnicity were controlled. It should be noted, however, that kindergarten SES was not a significant predictor of the shape, or total growth for the LRA group, although it was a significant predictor for the RD group. As described above, individual SES and school SES, although related, may both influence reading achievement (Marks, 2015), but school SES was not specifically considered in the current study. Thus, if school and neighborhood environmental factors were included in the model, then such variables may have significantly predicted the total reading growth of the LRA group. Given the lack of data about such school related variables in the current study, these hypotheses about why children in the LRA experienced a decline in their reading growth at fifth grade should be interpreted with caution and should not be over-generalized. Nonetheless, one must still consider the effects of factors that exist at levels greater than the individual (e.g., microsystem, mesosystem, exosystem, macrosystem) when thinking about why the RD group and LRA groups displayed different reading achievement trajectories after the spring of first grade.

Sensitivity Analyses

The results of the sensitivity analyses for the shape group comparison were similar to the results using the original groups. In other words, defining the LRA group using a third-grade only reading achievement score below the 25th percentile, a RD group that included those identified for special education services in kindergarten, a RD group that excluded children identified for special education services in fifth grade, and a RD group that excluded children identified for special education services in fourth and fifth grade did not alter the results. Therefore, the RD and LRA groups demonstrated different patterns of reading achievement from kindergarten through eighth grade even when different ways of defining the groups were used.

Perhaps most notable about the sensitivity analyses for the shape multiple group LGM was that even when children who were identified for special education services due to RD in fourth and fifth grade were excluded from the sample, results were the same. Those excluded for being diagnosed with RD after third or fourth grade are usually described as having late-emerging RD. Children with late-emerging LD tend to have reading problems that do not become noticeable until later elementary school grades, and they tend to have minimal reading problems in early school grades (Catts, Compton, Tomblin, & Bridges, 2012). That is, research suggests that children with late-emerging RD generally do not have early reading problems that are missed as a result of poor identification practices; rather, their reading problems truly do not appear until later ages (Catts et al., 2012; Leach et al., 2003). Children with late-emerging RD, however, may have early oral-language deficits that precede their reading problems (Catts et al., 2012). Catts et al. (2012) found that in their sample of 493 children, 42% of children classified as having RD were considered late-emergers, and that 52% of these late-emergers had reading comprehension problems, 36% had problems in word reading, and 12% had a combination of both comprehension and word reading problems. Thus, it appears that the majority of those with late-emerging RD tend to have more deficits in reading comprehension. In the current study, even when these children with late-emerging RD, who likely had more comprehension-based deficits as opposed to phonological awareness deficits, were excluded, the pattern of reading achievement of the RD group remained relatively unchanged.

Group Comparison of Level of Reading Achievement at Eighth Grade

The third and final important finding from the unweighted multiple group LGM indicated significant group differences in the reading achievement scores at the end of eighth grade. Based on this analysis, in the spring of eighth grade, the RD group significantly outperformed the LRA group. These results, however, must be interpreted with caution because the unconstrained model in this analysis produced negative intercept and slope variances, which suggests that the model may not have been the best fit for these data, so the results may not be entirely accurate. Nonetheless, these results should also be interpreted in light of the fact that a Wald test that compared the mean eighth-grade reading achievement scores, in essence an independent samples t-test, found significant group differences.

First, it should be noted that the results of the current study conflict with the findings of Francis et al. (1996), who found no significant differences between RD and LRA groups at 16 years of age. Importantly, Francis et al. (1996) did not consider whether children in their sample received special education services, and they defined their groups using their own diagnostic criteria, so it is likely that some of the children in both groups received special education services. Additionally, Francis et al. (1996) did not use IRT scale scores; rather, they used Rasch scores, which may not be the most appropriate way to measure change in achievement over time and could have contributed to the non-significant group differences at age 16.

Second, O'Malley et al. (2002) also examined the longitudinal trends in reading achievement between RD and LRA groups, but their analyses stopped at around seven years of age. These researchers found that by age seven, there were no significant differences between the

RD and LRA groups in many of the reading-related processes (i.e., phonemic awareness, rapid naming, perceptual discrimination, spelling, word reading; O'Malley et al., 2002). Interestingly, the two groups did differ at age seven in visual-motor integration and letter-sound ability, in which the RD group outperformed the LRA group in both of areas (O'Malley et al., 2002). Thus, it is possible that if these researchers had examined a longer period of time, then either group differences in more of the reading-related process would have become apparent or group differences in letter-sound ability and visual-motor integration would have remained significant. Also, O'Malley et al. (2002) did not examine differences in any reading comprehension related skills, which in the current study was the focus of the assessment by the eighth-grade direct reading assessment and may represent where most of the differences between the RD and LRA groups lie. Lastly, O'Malley et al. (2002) did not consider whether children in their sample were receiving special education services, so how the services students received in school influenced their reading growth over time is unknown. Overall, it appears that the present study is the first study to compare eighth-grade reading outcomes for students with reading problems based on whether they received special education services during their education as part of a longitudinal study. Accordingly, how special education services and instruction for students in each of these groups influenced their reading development over time must be discussed.

Instructional Considerations

Even though the results related to this research question are questionable, it is necessary to discuss what it means that the RD group performed significantly higher than the LRA group at the end of eighth grade. In thinking about why the RD group performed better than LRA group in the spring of eighth grade, one must consider the services and supports students in each of these groups received during their education. For both groups, however, exactly what their educational environments, instruction/intervention, and supplemental supports and services (e.g., accommodations) looked like is unknown, as mentioned above. Despite the uncertainty about their exact services, the discussions below will explain the current results in relation to what their classroom instruction/intervention and special services may have looked like.

RD and special education services. Generally, research suggests that children provided with special education services have lower academic achievement than students in general education settings and do not catch up to their non-disabled peers (Morgan et al., 2010, 2011) as well as that children in special education receive lower quality and less intensive instruction than children in general education classrooms (Bentum & Aaron, 2003; Dworet, 1987; Moody et al., 2000; Morsink et al., 1986; Vaughn et al., 2002). Interestingly, in the current study, the RD group's higher reading performance at the end of eighth grade provide some indication that special education services (e.g., supplemental instruction, classroom accommodations) were effective, relative to children with reading problems who did not receive special education services. Although special education is not the only factor that explains why children with RD performed better than the children with LRA in eighth grade, such that the groups were not exactly equal at kindergarten entry as well as that characteristics of their school and classroom instructional environments were unknown, it is plausible to suggest that there was something about the special education services from which the children in the RD group benefited.

While children in the RD group may have benefited from special education services in comparison to children with LRA who did not receive such services, it is likely that the RD group did not make enough progress over time to catch up to their peers without reading problems. In support of this idea, Morgan et al. (2010) found that children receiving special education services in third grade demonstrated significantly lower reading skills in fifth grade than closely matched peers not receiving special education services. Despite the lower scores in fifth grade, the two groups made statistically equivalent gains in reading skills between third and fifth grade (Morgan et al., 2010). Although Morgan et al. (2010) included children with all types of disabilities in the special education group, the results of that study are likely consistent with the results of the current study that suggest children with RD who received special education services do benefit from receipt of such services, but such services may not be enough to help those with RD to achieve at the same levels as their peers without reading impairments. Whether children in the RD group in the current sample did in fact display lower reading skills at the end of eighth grade compared to their non-impaired peers is unknown.

Still, one must consider that RD is a persistent, life-long learning disability. As a result, it may not be that special education is ineffective in remediating the reading problems of children with RD, but rather, children with RD are in general unlikely to catch up to their non-impaired peers, even with special accommodations. Research suggests that children with RD make significant gains in their reading development over time, yet they never make enough gains to close the gap between themselves and children without reading problems (Francis et al., 1996). When scientifically-based, intensive, and high-quality reading interventions are provided to students with RD, especially older children between 8 and 10 years of age, they may improve their reading accuracy and comprehension to fall within the average range, but they continue to demonstrate oral reading fluency rates that are severely impaired (Torgesen et al., 2001). It is important to note that these intensive reading interventions were not provided within the public-school system; rather, they were delivered as part of the research study, and such positive results may be even less likely within the school public school setting. This idea that children with RD will never close the gap between themselves and children with reading problems does not mean that it is fruitless to provide them with intensive intervention, special services, and accommodations to help them improve, as evidenced by the facts that (1) children with RD do make do make improvements in their reading problems and respond to intervention (Denton et al., 2006; Lovett et al., 1994; Torgesen, 2001, 2002; Vaughn et al., 2012), and (2) children with RD are better able to learn reading comprehension skills and access more advanced content when provided with accommodations when compared to children with LRA who likely did not receive such accommodations. Nonetheless, it is helpful to understand that children with RD will struggle with learning as a result of their disability throughout their schooling.

LRA and classroom instruction/intervention. During the years in which the data for the ECLS-K were collected, RTI frameworks were only beginning to be implemented in schools across the nation. Prior to RTI implementation, students typically did not receive early intensive reading intervention within the general education setting. Title I programs were one way in which children would be provided with supplemental reading instruction and support, if they were not receiving special education services; however, Title I programs have historically consisted of disjointed instruction that does not always align with the general curriculum (Johnston et al., 1985).

It was not until the late 1990s and early 2000s when researchers and educators began to think differently about the importance of evidence-based, high-quality reading instruction. During this time, questions were raised not only about how schools identify students as having a RD but also about how students should be taught how to read. Large scale research projects such as the one conducted by the National Reading Panel in 2000 were commissioned to investigate areas to focus reading instruction and the most effective ways to teach students to read. Generally, research suggests intervening in earlier grades (i.e., kindergarten, first grade, and second grade) leads to better results than does intervention in later grades, especially as a result of the intensity and duration needed when intervening at later grades (Foorman, Breier, & Fletcher, 2003). Additionally, small-group interventions have been found to be just as effective as one-on-one interventions, and well-trained paraprofessionals can implement interventions just as effectively as teachers (Foorman et al., 2003). Moreover, the most effective reading interventions are the ones that include explicit instruction in phonemic awareness, the alphabetic principle (i.e., grapheme-phoneme correspondence), and phonics instruction integrated with instruction in reading comprehension and opportunities to read and write based on subject content (Foorman et al., 2003). Also, as described in a previous section, it is critical that reading instruction is individualized and differentiated, such that students will respond differently to reading instruction depending on their reading skills prior to receiving said instruction (Connor et al., 2004a, 2004b; 2009). In the current study, although the type of instruction/intervention students in the LRA group received is unknown, it may be that if they had received more intensive, evidence-based instruction and intervention in reading especially at earlier ages as well as accommodations and interventions that continued to support their reading comprehension development as they got older, then they may have not fallen as far behind the RD group.

Sensitivity Analyses

The results of the sensitivity analyses for the eighth-grade group comparisons were partially similar to the results using the original groups. To start, when the RD group included those identified for special education services in kindergarten and when the RD group excluded children identified for special education services in fifth grade, the results did not differ from those with the original groups. However, when the LRA group was defined using a third-grade only reading achievement score below the 25th percentile and when the RD group excluded children identified for special education services in fourth and fifth grade, the results conflicted with those of the original groups, such that no significant group differences in reading scores were found at the end of eighth grade.

First, regarding the LRA group defined using a third-grade only reading achievement score below the 25^{th} percentile, the conflicting results must be interpreted with caution for two reasons. One reason is that models used in this analysis produced negative intercept and slope variances, which could have led to inaccurate results. The second reason is that the graph for this sensitivity analysis in Figure 7 shows a substantial gap between the two groups reading trajectories at the eighth-grade time point, perhaps one that is even bigger than in the graphs using the original groupings (see Figure 6). Thus, it is possible the negative variances in this multiple group LGM inaccurately led to results that suggest no group differences existed, when really group differences did exist.

Second, when the RD group excluded children identified for special education services in fourth and fifth grade, no significant group differences between the RD and LRA groups' reading achievement in eighth grade were identified. One possible explanation for this finding is that by excluding children with late-emerging RD (i.e., those identified after fourth grade; see discussion above), which was perhaps students with less severe reading problems who would have scored higher on the direct reading assessment, the remaining children in the RD group had lower scores on the direct reading assessment, performing more similarly to those in the LRA group. In other words, the RD group in this sensitivity analysis only included children with the most severe and impairing reading difficulties, causing them to perform worse on the direct reading assessment, which produced scores more similar to those in the LRA group. Further research is needed to truly understand how the RD group differs when students identified to receive special education services in fourth and fifth grade are included in the sample versus when they are excluded and how this affects their reading achievement trajectories and reading performance in eighth grade.

Implications for Practice

The current study has practical implications related to the identification of students with reading problems as well as prevention and intervention efforts. Perhaps, one of the most important implications concerns how schools identify students to receive special education services due to reading problems. The current study found no significant differences in the reading achievement between the RD and LRA group at kindergarten entry, but that the two groups go on to develop reading skills differently and have different outcomes with the RD group coming out with the advantage at the end of eighth grade. The lack of group differences in kindergarten underscores the need for (1) universal screenings as early as kindergarten in order to identify students at-risk for reading failure, (2) early intensive, evidence-based reading intervention for those identified at-risk and continued monitoring of their progress, and (3) a RTI approach to RD identification for students who do not respond to intensive reading interventions.

Universal reading screeners must measure the most important predictors of later reading achievement and be sensitive enough to identify students who are struggling with early reading processes and are not making appropriate growth. Universal reading screeners should measure not only phonemic and phonological awareness (e.g., letter naming fluency, letter sound fluency, segmentation, blending, nonsense word fluency) but also rapid naming, as research suggests that rapid naming is a key skill in predicting later reading difficulties (Boscardin, Muthén, Francis, & Baker, 2008; Schatschneider, Carlson, Francis, Foorman, & Fletcher, 2004; Wolf & Bowers, 2000) and may in fact differentiate those with a more severe double-deficit (Wolf & Bowers, 2000). Universal screeners that assess students early reading skills (e.g., letter naming fluency, letter sound fluency, segmentation, blending, nonsense word fluency) are typically only administered in kindergarten and early in first grade, and by the winter of first grade, universal screenings assess students' oral reading fluency. However, further research is needed to determine if early reading processes (e.g., phonological awareness) should be assessed through second or third grade when children are still learning to read in order to effectively identify students who have not yet mastered basic reading skills and are in need of more intensive intervention in these areas, as opposed to focusing solely on reading fluency interventions.

Once students are identified as at-risk for reading failure, intensive, empirically-supported, high-quality, and individualized instruction and intervention should be implemented immediately (Foorman et al., 2003). First, it is important that classroom instruction at the tier 1 level consist of research-based practices that can be differentiated to best meet students unique learning needs (Foorman et al., 2003; National Reading Panel, 2000; Connor et al., 2004a, 2004b, 2009). Second, those students identified at risk for reading failure (i.e., tier 2, tier 3) should be provided high-quality, evidence-based intervention designed to meet their specific reading deficits, and their progress with such interventions should be monitored frequently (e.g., weekly; Fletcher et al., 2004). Research supports the effectiveness of early reading intervention for student at-risk for reading problems, such that after provided with intensive high-quality intervention in kindergarten and/or first grade, for example, Vellutino et al. (2006) found that 84% of the at-risk children became average readers by first grade, and 58% of children who were at-risk until the beginning of first grade performed in the average range on all reading outcome measures at the end of first, second, and third grades. Thus, early intensive reading intervention is critical in preventing later reading failure, yet continued monitoring of these at-risk students' progress is also crucial to make sure that they continue to make progress, meet benchmarks, and do not fall further behind.

Although early intensive reading intervention is effective in preventing reading failure for many students, some children will not respond to reading intervention (e.g., close the gap between themselves and average readers), which is when a special education evaluation and subsequent special education placement may be warranted (Fletcher et al., 2004). Students in special education should receive intensive tier 2 and tier 3 intervention in addition to supplemental instruction and intervention in special education. Furthermore, special education supports for students with difficult to remediate reading problems should also continue to emphasize accommodations (e.g., read aloud, text to speech, keyboarding) to help these students learn and be successful in other subject areas in spite of their reading challenges.

Related to the current study, students were identified to receive special education likely based on an IQ-achievement discrepancy. However, if an RTI model had been used and only students who did not respond to intensive reading intervention were identified for special education, then perhaps the results would have been different, such that students with LRA would have received early intensive intervention in early elementary school grades and may have outperformed the RD group in later grades. Hence, using an RTI framework, one may expect that students in an LRA group would make significant growth in early elementary school grades as a result of intensive reading instruction and intervention that would allow them to catch up to their same-grade peers. If students with RD are those with the more severe deficits, who do not respond to intervention, then perhaps, one would continue to expect them to less frequently catch up to their same-grade peers, as suggested in previous research (Francis et al., 1996).

In addition to early prevention and intervention efforts, this study supports the need for general education accommodations for students who are identified as needing more intensive reading instruction/intervention in early school grades as a result of deficits in phonological awareness, decoding, and/or fluency. In the current study, students in the RD group likely received accommodations to help them continue to progress in their learning both in reading comprehension and in other subject areas in spite of their early reading difficulties (e.g., phonological awareness). Accordingly, when students are identified as in need of tier 2 or tier 3 reading supports via universal screenings, but perhaps are not yet in need of special education services, they should be eligible to receive accommodations that would help them continue to progress in their learning. Providing accommodations will help students with reading difficulties to not be halted by their challenges, and they will be able to continue to develop important reading related skills like vocabulary, critical thinking, and analysis. Accommodations can include read aloud support on assignments and tests, shortened assignments, opportunities to use keyboards to help with writing, etc. Whether these accommodations should be written into a more formalized plan, such as a Section 504 Plan, or should be left up for teachers to informally implement are important options for policy makers, educators, administrators, and researchers to consider, as they think about preventing students with LRA from falling further and further behind their peers over time.

Along with the need for general education accommodations for struggling readers, this study provides support for the need of continued universal screenings, intensive instruction and intervention related to reading comprehension. In the current study, as students got older, the direct reading assessment focused more on comprehension (e.g., vocabulary, understanding, personal reflection, critical stance), which suggests that reading growth and reading problems may be related to comprehension skills in addition to early reading skills, like decoding and fluency. Therefore, universal screenings of reading comprehension should occur in both earlier school grades and should persist throughout middle school. Schools would benefit from continued universal screening in reading comprehension as students get older in order to identify both when students' early reading skill deficits begin to influence their reading comprehension as well as when students who did not have early reading skill deficits begin to develop reading comprehension difficulties.

Furthermore, previous research indicates that different reading skills and variables predict reading problems at second grade versus eighth grade (Adolf et al., 2010). As such, when students reach later elementary and middle school grades, reading instruction should include research-based reading comprehension instruction and intervention. Some researchers have found that specific reading comprehension instructional strategies are most effective in improving reading comprehension skills and subsequent test scores, including individualized schema-based learning, conceptual learning, and transactional learning (Block, Parris, Reed, Whiteley, & Cleveland, 2009). Individualized schema-based learning is when students engage in independent silent reading with teachers providing scaffolding and support when students encounter barriers that prevent them from understanding the text (Block et al., 2009). Conceptual learning is based on the idea that reading comprehension increases when there are opportunities to engage in reading activities related to learning new knowledge about a topic (Block et al., 2009). Transactional learning involves both reading books independently and then engaging in group discussions with other students and the teacher about the book in order to deepen one's understanding of the text and make connections to their own lives and larger subject-related themes (Block et al., 2009). Both educators and researchers must continue to identify the most effective ways to target students' reading comprehension deficits as they get older and higher level reading comprehension skills that are need to be successful in school.

Both the RD and LRA groups experienced declines in their reading growth, albeit at different time points. The RD group experienced slower growth between the spring of third grade and the spring of fifth grade, and the LRA group experience slower growth from the spring of third grade to the spring of fifth grade and significant deceleration from the spring of fifth grade to the spring of eighth grade. Although it is impossible to determine whether these declines resulted from poor early reading skills that led to comprehension deficits, because children were unable to decode and fluently read text, or from comprehension deficits that developed from increasing school demands and the switch from "learning to read" to "reading to learn," one must consider how schools can best support these students learning in all subjects if they do not have grade-appropriate reading skills and do not respond/do not receive intervention. In other words, schools should consider putting accommodations in place for students with reading problems identified via universal screenings (described above), so that they do not lose motivation, have poor achievement in other areas, and develop more serious academic and socio-emotional problems. Although providing accommodations to all students with LRA does not recognize a larger systemic issue with reading instruction, the idea is that when a child is past the point of early reading intervention and current interventions are not helping the student to grow as quickly as desirable, educators should consider how they can best prevent further academic problems and school failure.

Limitations and Future Research

There are several limitations of this study. First, due to its non-experimental nature, no causal inferences can be drawn about the reasons for differences and similarities in the reading achievement of the two groups. That is, an experimental design would be necessary to establish the extent to which special education services influenced the development of participants' reading achievement over time. Future researchers should consider using more rigorous matching methods (e.g., propensity score matching) to ensure that the groups are more equivalent on variables aside from the services they received, because a true experimental study that utilizes random assignment cannot be conducted for this population (i.e., cannot randomly assign students to receive special education services). This study only provides a preliminary idea about the differences in the reading achievement trajectories between those with reading problems who did and did not receive special education services.

Second, the present study did not include a group of non-impaired students. As a result, it is not possible to examine how the RD and LRA groups compared to children with average reading skills, which limits the conclusions one can draw about a Matthew effect or a compensatory trajectory of reading development in this study. Consequently, this study would benefit from replication with the inclusion of a non-impaired, average-performing group in order to draw stronger, more meaningful conclusions.

Third, this study did not account for school- or classroom-level variables (e.g., school SES, instructional practices). It is likely that school- and classroom-level variables did play a role in students' reading achievement over time, but because these variables were not considered in the current study, how such variables played a role is unknown. Future research should address this problem by utilizing a multi-level latent growth model, in which differences in reading achievement resulting from school and classroom variables can be accounted for (see Muthén, 1997). Further, there was likely great heterogeneity in terms of what special education services looked like for those in the RD group as well as what supports were available for those in the

LRA group. In other words, this study did not permit for the examination or control of the type of supports and services the students may have received in school to help with their reading difficulties. As a result, researchers should conduct studies that include data on the specific type of reading instruction, reading intervention, and special education services students received.

Fourth, in addition to considering school- and classroom-related variables in future research, additional variables that were not included in the current study should considered. Some variables, such as IQ and language ability, were not included in the current sample because the dataset did not include such information, while other variables such as preschool experiences, home-literacy environment, and motivation were not included in order to maintain a parsimonious model. Nevertheless, researchers should consider including these variables in their work, since they have been shown to play a role in reading development over time.

Fifth, it is impossible to know which model of learning disability identification was used to diagnose students (e.g., discrepancy model, RTI) as well as why students in the LRA group were not identified to receive special education services. However, the purpose of this study was to examine the growth trajectories and trends in reading achievement between students identified with RD and students with LRA who were not identified, not necessarily to examine the validity of the models used to diagnose RD. Nonetheless, replicating this study using data that both indicates how students were identified to receive services and more current data that reflects RTI approaches to RD identification would allow researchers to better understand how students were identified to receive services, if they responded to previous intervention, and draw stronger conclusions about the effects of such services on students' reading growth trajectories.

Sixth, further research must consider alternate ways of defining the groups. In the current study, students in the RD group were included if they received special education services at any point between first and fifth grade. One way to improve this group definition would be to examine at what grade students were first identified and for how long they received services. Additionally, given that when the RD group excluded students identified in fourth and fifth grade, there were no differences between the RD and LRA groups' reading achievement in eighth grade, future research should consider defining the RD group based on their specific deficits. For example, one RD group might consist of students with only decoding deficits, while another group may consist of those with both decoding and comprehension deficits, and a third group might consist of students with only comprehension deficits. Similarly, for the LRA group, future research should seek to better examine when these students first experienced reading difficulties and in what specific areas they struggled. For both groups, it is plausible that the reading achievement trajectories would differ based on when children first developed reading problems/were first identified to receive services and where their specific reading deficits lie. Therefore, future research should utilize measures of reading achievement that are specific to individual reading skills (e.g., phonemic awareness, rapid naming, decoding, comprehension) to more closely examine the longitudinal differences in these skills between those with RD and LRA.

Seventh, the current study is limited in its generalizability in multiple ways. First, this study was not able to employ the weighted data as was originally intended. When the weighted data were used, the models frequently produced negative slope and intercept variances. Although this does not automatically rule out the use of the weighted data, it does suggest that the models with the weighted data were not the most appropriate models. Even though weighted model were not fully interpreted, the final unweighted sample was not drastically different from the weighted sample, in terms of gender, SES, race/ethnicity, age, and mean reading IRT scale scores, suggesting that the unweighted models may be somewhat generalizable. Also, conducting analyses using weighted data in MPlus does not provide certain fit indices, such as the chi-square statistics, CFI, and TLI, which is why examining both weighted and unweighted fit indices was important. However, it is still important that future research makes use of weighted data in order to generalize the results to the national population, adjust for differential selection probabilities, and reduce non-response bias. Second, the current study is limited in its generalizability because only those who had both sample weights and data present in the field management system (FMS) variable that identified participants as having/having not received special education services were included in the sample. In other words, those missing a sample weight and those missing data on the variable that identified them as having or having not received special education services were excluded from the sample. Unfortunately, this is a limitation of using an extant dataset. Lastly, it should be noted that the current sample was never intended to be nationally representative of the entire kindergarten class of 1998-1999, but rather be nationally representative of those with reading problems in the kindergarten class of 1998-1999. Overall, the results of the present study's analyses should be interpreted with caution and should not be generalized to the entire population of interest. Instead, the results should be interpreted as preliminary evidence of the RD and LRA groups' reading growth trajectories from kindergarten through eighth grade within only the sample included in the current study.

Finally, given the unusable weighted data along with the basis coefficients for the LRA group that were greater than one, future research should consider other possible ways to the model the data. One possibility would be to include a quadratic factor in the model, although this likely would not improve model fit for a sample that only includes the RD and LRA groups. Adding a quadratic factor to the unweighted unconditional latent growth model in the current study, in fact, worsened model fit relative to model fit of the latent basis model (Quadratic model: χ^2 = 42.16, df = 7, RMSEA = 0.07, CFI = 0.98, TLI = 0.95; Latent basis model: χ^2 = 10.52, df = 7, RMSEA = 0.02, CFI = 0.998, TLI = 0.995). Similarly, model fit was not improved by adding a quadratic to the weighted unconditional model, although the slope variance was positive (Quadratic model: RMSEA = 0.09; Latent basis model: RMSEA = 0.06). While a quadratic model did not appear to be better model for the given data, it is still important to consider, especially when a non-impaired group of students is added into the sample, as research tends to support the idea of quadratic growth in reading achievement within the general population (Francis et al., 1996).

A second possible way to model the data would be to utilize growth mixture modeling (GMM; Muthén, 2000; Muthén, Khoo, Francis, & Boscardin, 2002; Muthén & Muthén, 2000). GMM combines growth curve and latent class analysis, which allows growth trajectories to vary across different classes (Wu et al., 2014). Thus, this type of modeling identifies distinct groups of growth trajectories over time, which is advantageous because it does not require the model to fit to pre-determined groups, but it allows the data to classify individuals based on the data and growth trajectories. GMM has been used in other research that examines achievement growth trajectories (e.g., Boscardin et al., 2008; Wu et al., 2014). Given that the basis coefficients in the conditional latent basis model for the LRA group were greater than one, this suggests that the conditional latent basis model was not the best fit for the data. Using GMM may help solve this issue by finding the latent classes within the growth trajectories. If GMM were used, researchers must still consider whether students did or did not receive special education services, and it would be interesting to examine the relationship among the latent classes, their growth trajectories, whether students did or did not receive special education services, and the type of reading instruction and interventions students received.

Conclusions and Future Directions

This study provided an initial investigation into the reading growth trajectories of children with RD who received special education services and children with LRA who did not receive special education services. Limited prior research has examined differences in reading growth trajectories of children who were identified based on school eligibility using advanced statistical methods, such as non-linear LGM, over longer developmental periods. Thus, the current study makes a contribution to the literature by focusing on reading development during an extended period of time (i.e., kindergarten through eighth grade), by utilizing a non-linear latent growth model, and by relying on school-based special education eligibility to identify students in each group.

The findings of the current study suggest that children with reading difficulties experienced reading growth in a non-linear fashion. That is, children experienced the most substantial growth in their reading skills between the spring of first grade and the spring of third grade, which was followed by a deceleration in reading growth from the spring of third grade to the spring of the eighth grade. These results support other research that provides evidence of a decline in reading growth after the early elementary school years. As LGM techniques continue to advance, researchers should further study the patterns of reading growth over time, especially of those with reading deficits, using advanced models like S-shaped curves (e.g., Gompertz, Richards, Logistic) and piecewise models to better understand how reading growth changes over time in relation to different developmental periods.

The results of the current study support the notion that children who went on to develop reading problems, but did not all receive equitable school services, displayed similar reading skills at the beginning of kindergarten. That is, regardless of whether children were diagnosed with RD and received special education services or were not diagnosed with RD but displayed LRA, they entered kindergarten with similar reading skills. Furthermore, the reading growth of these two groups looked relatively similar until the end of first grade, when the two groups' patterns of reading development diverged. After first grade, children in the RD group who received special education services continued to make substantial progress in their reading skill development, and they even experienced accelerated growth from the spring of fifth grade to the spring of eighth grade. The LRA group, however, experienced substantial growth from the spring of first grade to the spring of third grade but experienced a deceleration from the spring of third grade through the spring of eighth grade, with a prominent decline in reading growth from the spring of fifth grade to the spring eighth grade. By eighth grade, the LRA group performed significantly worse in terms of their reading achievement compared to the RD group.

These findings highlight a several important areas for future research. First, research is needed to better understand the growth trajectories of children with LRA using more advanced statistical methods such as GMM. Second, future research should examine how specific school-based factors, such as instruction, intervention, and special education services influence the growth trajectories of the two groups. Research should consider when students began receiving services, how long services were received, and what comprised the instruction, intervention, and special education services. Additionally, further neuropsychological studies can help clarify whether there are differences between the two groups across cognitive measures as well as reading-related skills while taking students' school-based eligibility into account. Lastly, future research is necessary to elucidate the relationship between the longitudinal growth trajectories of students receiving and not receiving special education and specific reading-related skills and sub-processes. For example, when the RD group experienced reading growth between fifth and eighth grade, did such growth occur in early phonological and decoding skills or in more advanced reading comprehension skills? Likewise, how does the growth in these reading-related skills and sub-processes differ between the RD and LRA groups over time? Thus, the results of the current study underscore the importance of further longitudinal research regarding these two groups and their reading development over time in order to inform both early prevention and intervention efforts and school-based identification processes.

Finally, the current study was positioned within a bioecological framework as well as a stage theory of reading development (Chall, 1983); however, one must recognize both the limitations of Chall's stage theory and the relevance of other theoretical perspectives. Although Chall's theory recognizes the importance of environmental factors that influence reading development, which is more deeply reflected in Bronfenbrenner's bioecological theory, Chall's theory assumes that reading development occurs within one's own individual head. However, one must consider that learning to read is not only a biological phenomenon, but it is also a social one. Theorists like Vygostky believe that all higher psychological functions originate from social interactions and the social context (Englert & Mariage, 2013). That is, language and reading development develop from children's interactions with others and through meditational tools and signs of culture (e.g., books, games, iPads; Englert & Mariage, 2013). Sociocultural theories of reading development also emphasize the importance of the apprenticeship process in learning to read, in which children are apprenticed into culturally-relevant, valued social activities by community members (Englert & Mariage, 2013). Based on this sociocultural perspective, reading instruction in schools should emphasize an apprenticeship model in which teachers provide high support in the early stages of learning and then gradually decrease this support as students' progress, individualized support based on individual needs, activities that allow for collaboration amongst students, and opportunities for students to practice engaging with language and literature (Englert & Mariage, 2013). The sociocultural perspective on reading development is especially important as researchers and educators seek to understand the ways in which sociocultural variables, such as race/ethnicity, SES, and even gender, influence how children learn to read and develop reading skills over the course of their schooling. Ultimately, for children with reading problems both diagnosed and undiagnosed, future research should consider the sociocultural model of reading development in order to develop and implement instruction and intervention that focuses on not only specific reading-related skills but also the larger cultural context in which students are embedded.

APPENDIX

MICHIGAN STATE
UNIVERSITY

Initial IRB Application Determination *Exempt*

August 9, 2016

To: Jodene Fine
440 Erickson Hall

Re: **IRB# x16-1014e** Category: Exempt 4
Approval Date: August 9, 2016

Title: Longitudinal trends in the academic achievement and socio-emotional functioning of children with and without learning disabilities

The Institutional Review Board has completed their review of your project. I am pleased to advise you that **your project has been deemed as exempt** in accordance with federal regulations.

The IRB has found that your research project meets the criteria for exempt status and the criteria for the protection of human subjects in exempt research. **Under our exempt policy the Principal Investigator assumes the responsibilities for the protection of human subjects** in this project as outlined in the assurance letter and exempt educational material. The IRB office has received your signed assurance for exempt research. A copy of this signed agreement is appended for your information and records.

Renewals: Exempt protocols do not need to be renewed. If the project is completed, please submit an ***Application for Permanent Closure***.

Revisions: Exempt protocols do not require revisions. However, if changes are made to a protocol that may no longer meet the exempt criteria, a new initial application will be required.

Problems: If issues should arise during the conduct of the research, such as unanticipated problems, adverse events, or any problem that may increase the risk to the human subjects and change the category of review, notify the IRB office promptly. Any complaints from participants regarding the risk and benefits of the project must be reported to the IRB.

Follow-up: If your exempt project is not completed and closed after three years, the IRB office will contact you regarding the status of the project and to verify that no changes have occurred that may affect exempt status.

Please use the IRB number listed above on any forms submitted which relate to this project, or on any correspondence with the IRB office.

Good luck in your research. If we can be of further assistance, please contact us at 517-355-2180 or via email at IRB@msu.edu. Thank you for your cooperation.

Sincerely,

Harry McGee, MPH
SIRB Chair

c: Danielle Wexler

Office of Regulatory Affairs
Human Research Protection Programs

Biomedical & Health Institutional Review Board (BIRB)

Community Research Institutional Review Board (CRIRB)

Social Science Behavioral/Education Institutional Review Board (SIRB)

Olds Hall
408 West Circle Drive, #207
East Lansing, MI 48824
(517) 355-2180
Fax: (517) 432-4503
Email: irb@msu.edu
www.hrpp.msu.edu

REFERENCES

REFERENCES

Aaron, P. G. (1997). The impending demise of the discrepancy formula. *Review of Educational Research*, *67*, 461-502. https://doi.org/10.3102/00346543067004461

Aarnoutse, C., & van Leeuwe, J. (2000). Development of poor and better readers during the elementary school. *Educational Research and Evaluation*, *6*, 251-278. https://doi.org/10.1076/1380-3611(200009)6:3;1-a;ft251

Aarnoutse, C., van Leeuwe, J., Voeten, M., & Oud, H. (2001). Development of decoding, reading comprehension, vocabulary and spelling during the elementary school years. *Reading and Writing*, *14*, 61-89. https://doi.org/10.1023/A:1008128417862

Acock, A. (2005). Growth curves and extensions using Mplus [PDF]. Retrieved from http://www.caldar.org/presentations/summer%20institute/2008/Day-2%20Aug%2014-2008/Track%201/Growth%20Curve%20Presentation.pdf

Acock, A. (2012). What to do about missing values. In H. Cooper (Ed.), *APA handbook of research methods in psychology* (Vol. 3) (pp. 27-50). Washington, D.C.: American Psychological Association (APA).

Adlof, S. M., Catts, H. W., & Lee, J. (2010). Kindergarten predictors of second versus eighth grade reading comprehension impairments. *Journal of Learning Disabilities*, *43*, 332-345. https://doi.org/10.1177/0022219410369067

Aikens, N. L., & Barbarin, O. (2008). Socioeconomic differences in reading trajectories: The contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology*, *100*, 235-251. https://doi.org/10.1037/0022-0663.100.2.235

Albrecht, S. F., Skiba, R. J., Losen, D. J., Chung, C., & Middelberg, L. (2012). Federal policy on disproportionality in special education: Is it moving us forward?. *Journal of Disability Policy Studies,* 23, 14-25. https://doi.org/10.1177/1044207311407917

Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention a review of the literature. *Remedial and Special Education*, *23*, 300-316. https://doi.org/10.1177/07419325020230050501

Allington, R. L., & McGill-Franzen, A. (1989). School response to reading failure: Instruction for Chapter 1 and special education students in grades two, four, and eight. *The Elementary School Journal*, *89*, 529-542. https://doi.org/10.1086/461590

Allison, P. D. (2002). *Quantitative applications in the social sciences: Missing data*. Thousand Oaks, CA: Sage Publications, Inc.

American Federation of Teachers. (1998). *Teaching reading is rocket science: What expert teachers of reading should know and be able to do*. Retrieved from http://www.aft.org/sites/default/files/reading_rocketscience_2004.pdf

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Anderson, K. G. (1997). Gender bias and special education referrals. *Annals of Dyslexia*, *47*, 151-162. https://doi.org/10.1007/s11881-997-0024-8

Ansari, A. (2017). The selection of preschool for immigrant and native-born Latino families in the united states. *Early Childhood Research Quarterly, 41*, 149-160. https://dx.doi.org/10.1016/j.ecresq.2017.07.002

Armbruster, B. B., Lehr, F., Osborn, J. M. (2010). *Put reading first: The research building blocks for teaching children to read: Kindergarten through grade 3*. National Institute for Literacy, U.S. Department of Education, & National Institute of Child Health and Human Development. DIANE Publishing. Retrieved from https://lincs.ed.gov/publications/pdf/PRFbooklet.pdf

Arnold, E. M., Goldston, D. B., Walsh, A. K., Reboussin, B. A., Daniel, S. S., Hickman, E., & Wood, F. B. (2005). Severity of emotional and behavioral problems among poor and typical readers. *Journal of Abnormal Child Psychology*, *33*, 205-217. https://doi.org/10.1007/s10802-005-1828-9

August, D., & Shanahan, T. (Eds.). (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth.* Mahwah, NJ: Erlbaum.

Aunola, K., Leskinen, E., Onatsu-Arvilommi, T., & Nurmi, J. E. (2002). Three methods for studying developmental change: A case of reading skills and self-concept. *British Journal of Educational Psychology*, *72*, 343-364. https://doi.org/10.1348/000709902320634447

Bach, S., Brandeis, D., Hofstetter, C., Martin, E., Richardson, U., & Brem, S. (2010). Early emergence of deviant frontal fMRI activity for phonological processes in poor beginning readers. *Neuroimage*, *53*, 682-693. https://doi.org/10.1016/j.neuroimage.2010.06.039

Barnard, W. M. (2004). Parent involvement in elementary school and educational attainment. *Children and Youth Services Review*, *26*, 39-62. https://doi.org/10.1016/j.childyouth.2003.11.002

Bast, J., & Reitsma, P. (1997). Mathew effects in reading: A comparison of latent growth curve models and simplex models with structured means. *Multivariate Behavioral Research*, *32*, 135-167. https://doi.org/10.1207/s15327906mbr3202_3

Bast, J., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew effects in reading: results from a Dutch Longitudinal study. *Developmental Psychology*, *34*, 1373-1399. https://doi.org/10.1037/0012-1649.34.6.1373

Bentum, K. E. (2003). Does reading instruction in learning disability resource rooms really work?: A longitudinal study. *Reading Psychology*, *24*, 361-382. https://doi.org/10.1080/02702710390227387

Berninger, V. W., Abbott, R. D., Vermeulen, K., Ogier, S., Brooksher, R., Zook, D., & Lemos, Z. (2002). Comparison of faster and slower responders to early intervention in reading: Differentiating features of their language profiles. *Learning Disability Quarterly*, *25*, 59-76. https://doi.org/10.2307/1511191

Blachman, B. A. (2000). Phonological awareness. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3; pp. 483–502). Mahwah, NJ: Erlbaum.

Blair, C., & Scott, K. G. (2002). Proportion of LD placements associated with low socioeconomic status evidence for a gradient?. *The Journal of Special Education*, *36*, 14-22. https://doi.org/10.1177/00224669020360010201

Blair, C., Granger, D. A., Willoughby, M., Mills-Koonce, R., Cox, M., Greenberg, M. T., ... & Fortunato, C. K. (2011). Salivary cortisol mediates effects of poverty and parenting on executive functions in early childhood. *Child Development*, *82*, 1970-1984. https://doi.org/10.1111/j.1467-8624.2011.01643.x

Blair, C., & Raver, C. C. (2012). Child development in the context of adversity: Experiential canalization of brain and behavior. *American Psychologist*, *67*, 309-318. https://doi.org/10.1037/a0027493

Block, C. C., Parris, S. R., Reed, K. L., Whiteley, C. S., & Cleveland, M. D. (2009). Instructional approaches that significantly increase reading comprehension. *Journal of Educational Psychology*, *101*, 262-281. https://doi.org/10.1037/a0014319

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

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, N.J: Wiley-Interscience.

Boscardin, C. K., Muthén, B., Francis, D. J., & Baker, E. L. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational Psychology*, *100*, 192-208. https://doi.org/10.1037/0022-0663.100.1.192

Bowers, G., & Wolf, M. (1993). Theoretical links among naming speed, precise timing mechanisms, and orthographic skill in dyslexia. *Reading and Writing, 5,* 69-85. https://doi.org/10.1007/bf01026919

Bronfenbrenner, U., Ceci, S. J. (1994). Nature-nurture reconceptualized in developmental perspective: A bioecological model. *Psychological Review*, *101*, 568-586. https://doi.org/10.1037//0033-295x.101.4.568

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

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner (Ed.) *Handbook of child development: Theoretical models of human development* (6th ed., pp. 793 – 828). Hoboken, NJ: Wiley.

Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *The Future of Children*, *7*, 55-71. https://doi.org/10.2307/1602387

Cameron, C. E., Grimm, K. J., Steele, J. S., Castro-Schilo, L., & Grissmer, D. W. (2015). Nonlinear Gompertz curve models of achievement gaps in mathematics and reading. *Journal of Educational Psychology*, *107*, 789-804. https://doi.org/10.1037/edu0000009

Case, L. P., Speece, D. L., & Molloy, D. E. (2003). The validity of a response-to-instruction paradigm to identify reading disabilities: A longitudinal analysis of individual differences and contextual factors. *School Psychology Review*, *32*, 557-582. Retrieved from https://eric.ed.gov/?id=EJ823575

Catts, H. W., Compton, D., Tomblin, J. B., & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology*, *104*, 166-181. https://doi.org/10.1037/a0025323

Catts, H. W., Hogan, T. P., & Fey, M. E. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *Journal of learning disabilities*, *36*, 151-164. https://doi.org/10.1177/00222194030360020

Chall, J. S. (1983). *Stages of reading development*. New York, NY: McGraw-Hill.

Chall, J. S., & Jacobs, V. A. (2003). The classic study on poor children's fourth-grade slump. *American Educator*, *27*1, 14-15. Retrieved from https://www.aft.org/periodical/american-educator/spring-2003/classic-study-poor-childrens-fourth-grade-slump

Chall, J. S., Jacobs, V. A., & Baldwin, L. E. (1990). *The reading crisis: Why poor children fall behind.* Cambridge, MA.: Harvard University Press.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*, 330-351. https://doi.org/10.1037/1082-989x.6.4.330

Coltheart, M. (2005). Modeling reading: The dual-route approach. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 6-23). Oxford: Blackwell.

Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education*, *4*, 5-17. https://10.1080/13603110600574322

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100*, 589-608. https://doi.org/10.1037/0033-295x.100.4.589

Compton, D. L., Fuchs, D., Fuchs, L. S., Elleman, A. M., & Gilbert, J. K. (2008). Tracking children who fly below the radar: Latent transition modeling of students with late-emerging reading disability. *Learning and Individual Differences*, *18*, 329-337. https://doi.org/10.1016/j.lindif.2008.04.003

Connor, C. M., Morrison, F. J., & Katch, L. E. (2004a). Beyond the reading wars: Exploring the effect of child-instruction interactions on growth in early reading. *Scientific Studies of Reading*, *8*, 305-336. https://doi.org/10.1207/s1532799xssr0804_1

Connor, C. M., Morrison, F. J., & Petrella, J. N. (2004b). Effective reading comprehension instruction: Examining child x instruction interactions. *Journal of Educational Psychology*, *96*, 682-698. https://doi.org/10.1037/0022-0663.96.4.682

Connor, C. M., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., ... & Morrison, F. J. (2009). Individualizing student instruction precisely: Effects of child× instruction interactions on first graders' literacy development. *Child Development*, *80*, 77-100. https://doi.org/10.1111/j.1467-8624.2008.01247.x

Cosden, M., Zimmer, J., & Tuss, P. (1993). The impact of age, sex, and ethnicity on kindergarten entry and retention decisions. *Educational Evaluation and Policy Analysis*, *15*, 209-222. https://doi.org/10.2307/1164422

Coutinho, M. J., & Oswald, D. P. (2000). Disproportionate representation in special education: A synthesis and recommendations. *Journal of Child and Family Studies*, *9*, 135-156. https://doi.org/10.1023/A:1009462820157

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: the indirect role of parental expectations and the home environment. *Journal of Family Psychology*, *19*, 294-304. https://doi.org/10.1037/0893-3200.19.2.294

Demir, Ö. E., & Küntay, A. C. (2014). Cognitive and neural mechanisms underlying socioeconomic gradients in language development: New answers to old questions. *Child Development Perspectives*, *8*, 113-118. https://doi.org/10.1111/cdep.12069

Demirtas, H., Freels, S. A., & Yucel, R. M. (2008). Plausibility of multivariate normality assumption when multiply imputing non-Gaussian continuous outcomes: a simulation assessment. *Journal of Statistical Computation and Simulation*, *78*, 69-84. https://doi.org/10.1080/10629360600903866

Denton, C. A., Fletcher, J. M., Anthony, J. L., & Francis, D. J. (2006). An evaluation of intensive intervention for students with persistent reading difficulties. *Journal of Learning Disabilities*, *39*, 447-466. https://doi.org/10.1177/00222194060390050601

Denton, C. A., Tolar, T. D., Fletcher, J. M., Barth, A. E., Vaughn, S., & Francis, D. J. (2013). Effects of tier 3 intervention for students with persistent reading difficulties and characteristics of inadequate responders. *Journal of Educational Psychology*, *105*, 633-648. https://doi.org/10.1037/a0032581

Denton, K., & West, J. (2002). Children's reading and mathematics achievement in kindergarten and first grade (NCES Publication No. 2002-125). Washington, DC: U.S. Department of Education. Retrieved from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2002125

Dickinson, D. K., & McCabe, A. (2001). Bringing it all together: The multiple origins, skills, and environmental supports of early literacy. *Learning Disabilities Research and Practice*, *16*, 186-202. https://doi.org/10.1111/0938-8982.00019

Doehring, D. G. (1978). The tangled web of behavioral research on developmental dyslexia. In A. L. Benton & D. Pearl (Eds.), *Dyslexia* (pp. 123-135). New York, New York: Oxford University Press.

Donahoe, K., & Zigmond, N. (1990). Academic grades of ninth-grade urban learning-disabled students and low-achieving peers. *Exceptionality*, *1*, 17-27. https://doi.org/10.1080/09362839009524739

Dong, Y., & Peng, C. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, *2*, 1-17. https://doi.org/10.1186/2193-1801-2-222

Duncan, T. E., Duncan, S.C., Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications* (2nd ed.). Mahwah, NJ: L. Erlbaum Associates.

Dworet, D. (1987, May). The effect of teacher expectation on reading instruction in special education self-contained classrooms. Paper presented at the International Reading Association meeting, Anaheim, CA.

Education for All Handicapped Children Act of 1975, P.L. 94-142, 20 U.S.C. § 1400 *et seq*.

Englert. C.S., & Mariage, T. (2013) The sociocultural model as a framework in instructional intervention research. In H. L. Swanson, K. R., Harris, & S. Graham (Eds.), *Handbook of Learning Disabilities*. New York, NY: The Guilford Press.

Englund, M. M., Luckner, A. E., Whaley, G. J. L., & Egeland, B. (2004). Children's achievement in early elementary school: Longitudinal effects of parental involvement, expectations, and quality of assistance. *Journal of Educational Psychology*, *96*, 723-730. https://doi.org/10.1037/0022-0663.96.4.723

Etmanskie, J. M., Partanen, M., & Siegel, L. S. (2016). A longitudinal examination of the persistence of late emerging reading disabilities. *Journal of Learning Disabilities*, 49, 21-35. https://doi.org/10.1177/0022219414522706

Field, A. (2009). *Discovering statistics using SPSS* (4th ed.) Thousand Oaks, CA: Sage Publications, Inc.

Fletcher, J. M., Coulter, W. A., Reschly, D. J., & Vaughn, S. (2004). Alternative approaches to the definition and identification of learning disabilities: Some questions and answers. *Annals of Dyslexia*, *54*, 304-331. https://doi.org/10.1007/s11881-004-0015-y

Fletcher, J. M., Foorman, B. R., Shaywitz, S. E., & Shaywitz, B. A. (1999). Conceptual and methodological issues in dyslexia research: A lesson for developmental disorders. In H. Tager-Flusberg (Ed.), *Neurodevelopmental Disorders* (pp. 271-306). Cambridge, MA: MIT.

Fletcher, J. M., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Liberman, I. Y., Stuebing, K. K., ... & Shaywitz, B. A. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology*, *86*, 6-23. https://doi.org/10.1037/0022-0663.86.1.6

Fletcher, J.M., Stuebing, K.K., Barth, A.E., Denton, C.A., Cirino, P.T., Francis, D.J., & Vaughn, S. (2011). Cognitive correlates of inadequate response to reading intervention. *School Psychology Review*, *40*, 3-22. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23125475

Foorman, B. R., Chen, D. T., Carlson, C., Moats, L., Francis, D. J., & Fletcher, J. M. (2003). The necessity of the alphabetic principle to phonemic awareness instruction. *Reading and Writing*, *16*, 289-324. https://doi.org/10.1023/A:1023671702188

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 curves analysis. *Journal of Educational Psychology*, *88*, 3-17. https://doi.org/10.1037/0022-0663.88.1.3

Frijters, J. C., Lovett, M. W., Steinbach, K. A., Wolf, M., Sevcik, R. A., & Morris, R. D. (2011).

Neurocognitive predictors of reading outcomes for children with reading disabilities. *Journal of Learning Disabilities*, *44*, 150-166. https://doi.org/10.1177/0022219410391185

Fuchs, D., Fuchs, L., Mathes, P., Lipsey, M., & Roberts, H. (2002). Is "learning disabilities" just a fancy term for low achievement? A meta- analysis of reading differences between low achievers with and without the label. In R. Bradley, L. Danielson, & D. Hallahan (Eds.), *Learning disabilities: Research to practice* (pp. 737-762). Mahwah, NJ: Erlbaum.

Fuchs, D., & Young, C. L. (2006). On the irrelevance of intelligence in predicting responsiveness to reading instruction. *Exceptional Children*, *73*, 8-30. https://doi.org/10.1177/001440290607300101

Gaffert, P., Meinfelder, F., & Bosch, V. (2016). Towards an MI-proper predictive mean matching. Retrieved from https://www.uni-bamberg.de/fileadmin/uni/fakultaeten/sowi_lehrstuehle/statistik/Personen/Dateien_Florian/properPMM.pdf

Gallagher, A., Frith, U., & Snowling, M. J. (2000). Precursors of literacy delay among children at genetic risk of dyslexia. *Journal of Child Psychology and Psychiatry*, *41*, 203-213. https://doi.org/10.1017/s0021963099005284

Gamse, B. C., Bloom, H. S., Kemple, J. J., Jacob, R. T., (2008). *Reading First Impact Study: Interim Report* (NCEE 2008-4016). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from https://www.mdrc.org/sites/default/files/full_488.pdf

Garson, G. D. (2015). *Missing values analysis and data imputation*. Asheboro, NC: Statistical Publishing Associates.

Geva, E., & Massey-Garrison, A. (2013). A comparison of the language skills of ELLs and monolinguals who are poor decoders, poor comprehenders, or normal readers. *Journal of Learning Disabilities*, *46*, 387-401. https://doi.org/10.1177/0022219412466651

Goswami, U. (1993). Phonological skills and learning to read. *Annals of the New York Academy of Sciences*, *682*, 296-311. https://doi.ord/10.1111/j.1749-6632.1993.tb22977.x

Goyette, K., & Xie, Y. (1999). Educational expectations of Asian American youths: Determinants and ethnic differences. *Sociology of Education*, 22-36. https://doi.org/10.2307/2673184

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549-576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*, 206–213. https://doi.org/10.1007/s11121-007-0070-9

Gresham, F. M. (2007). Evolution of the Response-to-Intervention Concept: Empirical Foundations and Recent Developments. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention (*pp. 10-24). New York: Springer.

Gresham, F. M., & Vellutino, F. R. (2010). What is the role of intelligence in the identification of specific learning disabilities? Issues and clarifications. *Learning Disabilities Research & Practice*, *25*, 194-206. https://doi.org/10.1111/j.1540-5826.2010.00317.x

Grimm, K. J., Ram, N., & Estabrook, R. (2017). *Growth modeling: Structural equation and multilevel modeling approaches*. NY, NY: The Guilford Press.

Grimm, K. J., Ram, N., & Hamagami, F. (2011). Nonlinear growth curves in developmental research. *Child Development*, *82*, 1357-1371. https://doi.org/10.1111/j.1467-8624.2011.01630.x

Guthrie, J. T., Wigfield, A., Metsala, J. L., & Cox, K. E. (1999). Motivational and cognitive predictors of text comprehension and reading amount. *Scientific Studies of Reading*, *3*, 231-256. https://doi.org/10.1207/s1532799xssr0303_3

Hackman, D. A., Gallop, R., Evans, G. W., & Farah, M. J. (2015). Socioeconomic status and executive function: Developmental trajectories and mediation. *Developmental Science*, *18*, 686-702. https://doi.org/10.1111/desc.12246

Hale, J. B., Kaufman, A., Naglieri, J. A., & Kavale, K. A. (2006). Implementation of IDEA: Integrating response to intervention and cognitive assessment methods. *Psychology in the Schools*, *43*, 753-770. https://doi.org/10.1002/pits.20186

Hair, N. L., Hanson, J. L., Wolfe, B. L., & Pollak, S. D. (2015). Association of Child Poverty, Brain Development, and Academic Achievement. *JAMA Pediatrics*, *169*, 822-829. https://doi.org/10.1001/jamapediatrics.2015.1475

Haukoos, J. S., & Newgard, C. D. (2007). Advanced statistics: Missing data in clinical research—Part 1: An introduction and conceptual framework. *Academic Emergency Medicine*, *14*, 662-668. https://doi.org/10.1111/j.1553-2712.2007.tb01855.x

Haynes, M. C., & Jenkins, J. R. (1986). Reading instruction in special education resource rooms. *American Educational Research Journal*, *23*, 161-190. https://doi.org/10.2307/1162952

Heiervang, Anders Lund, Jim Stevens, E. (2001). Behaviour problems in children with dyslexia. *Nordic Journal of Psychiatry*, *55*, 251-256. https://dx.doi.org/10.1080/080394801681019101

Hibel, J., Farkas, G., & Morgan, P. L. (2010). Who is placed into special education?. *Sociology of Education*, *83*, 312-332. https://doi.org/10.1177/0038040710383518

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2007). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*, 172-177. https://doi.org/10.1111/j.1750-8606.2008.00061.x

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology*, *49*, 4-14. https://doi.org/10.1037/a0027238

Hoff, E., & Core, C. (2015). What clinicians need to know about bilingual development. *Seminars in Speech and Language, 36*, 89-99. https://dx.doi.org/10.1055/s-0035-1549104

Holm, A., Crosbie, S., & Dodd, B. (2007). Differentiating normal variability from inconsistency in children's speech: normative data. *International Journal of Language & Communication Disorders*, *42*, 467-486. https://doi.org/10.1080/13682820600988967

Hooper, D., Coughlan, J., & Mullen, M. (2008). *Evaluating model fit: a synthesis of the structural equation modelling literature*. In A. Brown (Ed.), *7th European Conference on research methodology for business and management studies.* Paper presented at the European Conference on Research Methods for Business and Management, London, UK.

Hoskyn, M., & Swanson, H. L. (2000). Cognitive processing of low achievers and children with reading disabilities: A selective meta-analytic review of the published literature. *School Psychology Review*, *29*, 102-119. https://doi.org/10.1177/073194871003300101

Hosp, J. L., & Reschly, D. J. (2003). Referral rates for intervention or assessment a meta-analysis of racial differences. *The Journal of Special Education*, *37*, 67-80. https://doi.org/10.1177/00224669030370020201

Huang, F. L., Moon, T. R., & Boren, R. (2014). Are the reading rich getting richer? Testing for the presence of the Matthew effect. *Reading & Writing Quarterly*, *30*, 95-115. https://doi.org/10.1080/10573569.2013.789784

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children’s language growth. *Cognitive Psychology*, *61*, 343-365. https://doi.org/10.1016/j.cogpsych.2010.08.002

IDEA Amendments of 1997, 20 U.S.C §1415.

Individuals with Disabilities Education Act of 1990, 20 U.S.C. § 1400 *et seq.*

Individuals with Disabilities Education Act of 2004, 20 U.S.C. § 1400

Jimerson, S., Egeland, B., & Teo, A. (1999). A longitudinal study of achievement trajectories:

Factors associated with change. *Journal of Educational Psychology*, *91*, 116-126. https://doi.org/10.1037/0022-0663.91.1.116

Johnson, D. R., & Young, R. (2011). Toward best practices in analyzing datasets with missing data: Comparisons and recommendations. *Journal of Marriage and Family*, *73*, 926-945. https://doi.org/10.1111/j.1741-3737.2011.00861.x

Johnson, S. B., Riis, J. L., & Noble, K. G. (2016). State of the art review: Poverty and the developing brain. *Pediatrics*, *137*. Advanced online publication. https://doi.org/10.1542/peds.2015-3075

Johnston, P., Allington, R., & Afflerbach, P. (1985). The congruence of classroom and remedial reading instruction. *The Elementary School Journal*, *85*, 465-477. https://doi.org/10.1086/461414

Jorm, A. F., Share, D. L., Maclean, R., & Matthews, R. (1986). Cognitive factors at school entry predictive of specific reading retardation and general reading backwardness: A research note. *Journal of Child Psychology and Psychiatry*, *27*, 45-54. https://doi.org/10.1111/j.1469-7610.1986.tb00620.x

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, *80*, 437-447. https://doi.org/10.1037/0022-0663.80.4.437

Katz, L. A., Stone, C. A., Carlisle, J. F., Corey, D. L., & Zeng, J. (2008). Initial progress of children identified with disabilities in Michigan's Reading First schools. *Exceptional Children*, *74*, 235-256. https://doi.org/10.1177/001440290807400206

Kavale, K. A., & Flanagan, D. P. Ability–Achievement Discrepancy, Response to Intervention, and Assessment of Cognitive Abilities/Processes in Specific Learning Disability Identification: Toward a Contemporary Operational Definition. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention (*pp. 130-147). New York: Springer.

Kieffer, M. J. (2008). Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learners in the United States. *Journal of Educational Psychology*, *100*, 851-868. https://doi.org/10.1037/0022-0663.100.4.851

Kieffer, M. J. (2010). Socioeconomic status, English proficiency, and late-emerging reading difficulties. *Educational Researcher*, *39*, 484-486. https://doi.org/10.3102/0013189x10378400

Kline, R. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford.

Korkmaz, S., Goksuluk, D., Zararsiz, G. (2014). MVN: An R Package for Assessing Multivariate Normality. *The R Journal*, *6,* 151-162. Retrieved from https://journal.r-project.org/archive/2014-2/korkmaz-goksuluk-zararsiz.pdf

Kreisman, D. (2012). The source of African American–White inequality in early language acquisition: Evidence from Early Head Start. *Social Science Research*, *41*, 1429-1450. https://doi.org/10.1016/j.ssresearch.2012.05.010

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*, *95*, 3-21. https://doi.org/10.1037/0022-0663.95.1.3

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293-323. https://doi.org/10.1598/0872075028.40

Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology*, *95*, 211-224. https://doi.org/10.1037/0022-0663.95.2.211

Learning First Alliance. (1988). *Every child reading*. Retrieved from http://www.learningfirst.org/sites/learningfirst/files/assets/EveryChildReading.pdf

Lee, J. (2010). Tripartite Growth Trajectories of reading and math achievement: Tracking national academic progress at primary, middle, and high school levels. *American Educational Research Journal*, *47*, 800-832. https://doi.org/10.3102/0002831210365009

Lee, K. J., & Carlin, J. B. (2010). Multiple imputation for missing data: Fully conditional specification versus multivariate normal imputation. *American Journal of Epidemiology*, *171*, 624-632. https://doi.org/10.1093/aje/kwp425

Lee, J., Grigg, W., & Donahue, P. (2007). The nation's report card: Reading 2007 (NCES Publication No. 2007-496). Washington, DC: National Center for Education. Retrieved from https://nces.ed.gov/nationsreportcard/pdf/main2007/2007496.pdf

Lepola, J., Poskiparta, E., Laakkonen, E., & Niemi, P. (2005). Development of and relationship between phonological and motivational processes and naming speed in predicting word recognition in grade 1. *Scientific Studies of Reading*, *9*, 367-399. https://doi.org/10.1207/s1532799xssr0904_3

Leppänen, U., Niemi, P., Aunola, K., & Nurmi J. E. (2004). Development of reading skills among preschool and primary school pupils. *Reading Research Quarterly*, *39*, 72-93. https://doi.org/10.1598/rrq.39.1.5

Lewis, C., Hitch, G. J., & Walker, P. (1994). The prevalence of specific arithmetic difficulties and specific reading difficulties in 9-to 10-year-old boys and girls. *Journal of Child Psychology and Psychiatry*, *35*, 283-292. https://doi.org/10.1111/j.1469-7610.1994.tb01162.x

Liberman, I. Y. (1971). Basic research in speech and lateralization of language: Some implications for reading disability. *Bulletin of the Orton Society*, 21, 71-87. https://doi.org/10.1007/bf02663712

Liberman, I. Y., & Shankweiler, D. (1979). Speech, the alphabet and teaching to read. In L. Resnick & P. Weaver (Eds.), *Theory and practice of early reading* (pp. 109-132). Hillsdale, NJ: Erlbaum.

Liberman, I. Y., & Shankweiler, D. (1991). Phonology and beginning reading — a tutorial. In L. Rieben & C. Perfetti (Eds.), *Learning to read: Basic research and its implications* (pp. 3-17). Hillsdale, NJ: Erlbaum.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, *18*, 201-212. https://doi.org/10.1016/0022-0965(74)90101-5

Lipka, O., Lesaux, N. K., & Siegel, L. S. (2006). Retrospective analyses of the reading development of grade 4 students with reading disabilities risk status and profiles over 5 years. *Journal of Learning Disabilities*, *39*, 364-378. https://doi.org/10.1177/00222194060390040901

Livingstone, M. S., Rosen, G. D., Drislane, F. W., & Galaburda, A. M. (1991). Physiological and anatomical evidence for a magnocellular defect in developmental dyslexia. *Proceedings of the National Academy of Sciences*, *88*, 7943-7947. https://doi.org/10.1073/pnas.90.6.2556e

Lovegrove, W. J., Bowling, A., Badcock, D., & Blackwood, M. (1980). Specific reading disability: differences in contrast sensitivity as a function of spatial frequency. *Science*, *210*, 439-440. https://doi.org/10.1126/science.7433985

Lovett, M. W., Borden, S. L., DeLuca, T., Lacerenza, L., Benson, N. J., & Brackstone, D. (1994). Treating the core deficits of developmental dyslexia: Evidence of transfer of learning after phonologically-and strategy-based reading training programs. *Developmental Psychology*, *30*, 805-822. https://doi.org/10.1037/0012-1649.30.6.805

Lyon, G. R. (1985). Identification and remediation of learning disability subtypes: Preliminary findings. *Learning Disabilities Focus, 1,* 21-25. Retrieved from http://psycnet.apa.org/record/1989-13644-001

Maki, K. E., Floyd, R. G., & Roberson, T. (2015). State learning disability eligibility criteria: A comprehensive review. *School Psychology Quarterly*, *30*, 457-469. https://doi.org/10.1037/spq0000109

Mann, V. A., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory. *Journal of Learning Disabilities*, *17*, 592-599. https://doi.org/10.1177/002221948401701005

Mann, E. A., McCartney, K., & Park, J. M. (2007). Preschool predictors of the need for early remedial and special education services. *The Elementary School Journal*, *107*, 273-285. https://doi.org/10.1086/511707

Marks, G. N. (2015). Are school-SES effects statistical artefacts? Evidence from longitudinal population data. *Oxford Review of Education*, *41*, 122-144. https://doi.org/10.1080/03054985.2015.1006613

Mathes, P. G., & Torgesen, J. K. (1998). All children can learn to read: Critical care for the prevention of reading failure. *Peabody Journal of Education*, *73*, 317-340. https://doi.org/10.1207/s15327930pje7303&4_15

McArdle, J. J., & Epstein, D. B. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *29*, 110-133. https://doi.org/10.2307/1130295

McCoach, D. B., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology*, *98*, 14-28. https://doi.org/10.1037/0022-0663.98.1.14

McKinney, J. D., & Feagans, L. (1984). Academic and behavioral characteristics of learning disabled children and average achievers: Longitudinal studies. *Learning Disability Quarterly*, *7*, 251-265. https://doi.org/10.2307/1510483

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children*, *71*, 445-463. https://doi.org/10.1177/001440290507100404

McNamara, J. K., Scissons, M., & Dahleu, J. (2005). A longitudinal study of early identification markers for children at-risk for reading disabilities: The Matthew effect and the challenge of over-identification. *Reading Improvement*, *42*, 80-97. https://doi.org/10.1177/0022219411410040

McNamara, J. K., Scissons, M., & Gutknecth, N. (2011). A longitudinal study of kindergarten children at risk for reading disabilities: The poor really are getting poorer. *Journal of Learning Disabilities*, *44*, 421-430. https://doi.org/10.1177/0022219411410040

Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: a meta-analytic review. *Psychological Bulletin, 138*, 322-352. https://doi.org/10.1037/a0026744.supp

Meredith, W. M., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107-122. https://doi.org/10.1007/bf02294746

Miles, S. B., & Stipek, D. (2006). Contemporaneous and longitudinal associations between social behavior and literacy achievement in a sample of low-income elementary school children. *Child Development*, *77*, 103-117. https://doi.org/10.1111/j.1467-8624.2006.00859.x

Miskel, C., & Song, M. (2004). Passing Reading First: Prominence and processes in an elite policy network. *Educational Evaluation and Policy Analysis*, *26*, 89-109. https://doi.org/10.3102/01623737026002089

Molfese, V. J., Modglin, A., & Molfese, D. L. (2003). The role of environment in the development of reading skills: A longitudinal study of preschool and school-age measures. *Journal of Learning Disabilities*, *36*, 59-67. https://doi.org/10.1177/00222194030360010701

Moody, S. W., Vaughn, S., Hughes, M. T., & Fischer, M. (2000). Reading instruction in the resource room: Set up for failure. *Exceptional Children*, *66*, 305-316. https://doi.org/10.1177/001440290006600302

Morgan, P. L., Farkas, G., Tufis, P. A., & Sperling, R. A. (2008a). Are reading and behavior problems risk factors for each other?. *Journal of Learning Disabilities*, *41*, 417-436. https://doi.org/10.1177/0022219408321123

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

Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: Who falls increasingly behind?. *Journal of Learning Disabilities*, *44*, 472-488. https://doi.org/10.1177/0022219411414010

Morgan, P. L., Frisco, M. L., Farkas, G., & Hibel, J. (2010). A propensity score matching analysis of the effects of special education services. *The Journal of Special Education*, 43, 236-254. https://doi.org/10.1177/0022466908323007

Morgan, P. L., & Fuchs, D. (2007). Is there a bidirectional relationship between children's reading skills and reading motivation?. *Exceptional Children*, *73*, 165-183. https://doi.org/10.1177/001440290707300203

Morgan, P. L., Fuchs, D., Compton, D. L., Cordray, D. S., & Fuchs, L. S. (2008c). Does early reading failure decrease children's reading motivation?. *Journal of Learning Disabilities*, *41*, 387-404. https://doi.org/10.1177/0022219408321112

Morsink, C. V., Soar, R. S., Soar, R. M., & Thomas, R. (1986). Research on teaching: Opening the door to special education classrooms. *Exceptional Children*, *53*, 32-40. https://doi.org/10.1177/001440298605300103

Morrison, F. J., Alberts, D. M., & Griffith, E. M. (1997). Nature–nurture in the classroom:

Entrance age, school readiness, and learning in children. *Developmental Psychology*, *33*, 254-262. https://doi.org/10.1037/0012-1649.33.2.254

Muthén, B. O. (1997). Latent variable growth modeling with multilevel data. In M. Berkane (Ed.), *Latent variable modeling and applications to causality* (pp. 149-161). New York, NY: Springer 149-161.

Muthén, B. O. (2000). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: New developments and techniques* (pp. 1-33). Mahwah, NJ: Erlbaum.

Muthén, B. O., Khoo, S. T., Francis, D., & Boscardin, C. (2002). Analysis of reading skills development from kindergarten through first grade: An application of growth mixture modeling to sequential processes. In S. R. Reise & N. Duan (Eds.), *Multilevel modeling: Methodological advances, issues, and applications* (pp. 71-89). Mahwah, NJ: Erlbaum.

Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24,* 882-891. https://doi.org/10.1097/00000374-200006000-00020

Muthén, B. O., & Muthén, L. K. (2012). Mplus Version 7: User's guide. *Los Angeles, CA: Muthén & Muthén.*

Najarian, M. Pollack, J.M., and Sorongon, A.G. (2009). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Psychometric Report for the Eighth Grade* (NCES 2009–002). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubs2009/2009002.pdf

Nakao, K., and Treas, J. (1992). The 1989 Socioeconomic Index of Occupations: Construction from the 1989 Occupational Prestige Scores: *General Social Survey Methodological Report No. 74.* Chicago: NORC. Retrieved from http://gss.norc.org/Documents/reports/methodological-reports/MR074.pdf

Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment?. *Journal of Speech, Language, and Hearing Research*, *47*, 199-211. https://doi.org/10.1044/1092-4388(2004/017)

National Center for Education Statistics (NCES). (2002). *The Condition of Education 2002.* Retrieved from https://nces.ed.gov/pubs2002/2002025.pdf

National Center for Education Statistics (NCES). (2016). *The Condition of Education 2016.* Retrieved from https://nces.ed.gov/pubs2016/2016144.pdf

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the* *scientific literature on reading and its implications for reading instruction.* Washington, DC: National Institute of Child Health and Human Development. Retrieved from https://www.nichd.nih.gov/publications/pubs/nrp/Documents/report.pdf

Nelson, J. R., Benner, G. J., & Gonzalez, J. (2003). Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta-analytic review. *Learning Disabilities Research & Practice*, *18*, 255-267. https://doi.org/10.1111/1540-5826.00080

Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental Science*, *8*, 74-87. https://doi.org/10.1111/j.1467-7687.2005.00394.x

No Child Left Behind Act of 2002, Public Law No. 107–110, 115 Stat. 1425, U.S.C.

O'Connor, R. E., & Klingner, J. (2010). Poor Responders in RTI. *Theory Into Practice*, *49*, 297-304. https://doi.org/10.1080/00405841.2010.510758

Olson, R. K. (2002). Dyslexia: nature and nurture. *Dyslexia*, *8*, 143-159. https://doi.org/10.1002/dys.228

O'Malley, K. J., Francis, D. J., Foorman, B. R., Fletcher, J. M., & Swank, P. R. (2002). Growth in precursor and reading-related skills: Do low-achieving and IQ-discrepant readers develop differently?. *Learning Disabilities Research & Practice*, *17*, 19-34. https://doi.org/10.1111/1540-5826.00029

Oswald, D. P., Best, A. M., Coutinho, M. J., & Nagle, H. A. (2003). Trends in the special education identification rates of boys and girls: A call for research and change. *Exceptionality*, *11*, 223-237. https://doi.org/10.1207/s15327035ex1104_3

Oswald, D. P., Coutinho, M. J., Best, A. M., & Singh, N. N. (1999). Ethnic representation in special education the influence of school-related economic and demographic variables. *The Journal of Special Education*, *32*, 194-206. https://doi.org/10.1177/002246699903200401

Parrila, R., Aunola, K., Leskinen, E., Nurmi, J.-E., & Kirby, J. R. (2005). Development of individual differences in reading: Results from longitudinal studies in English and Finnish. *Journal of Educational Psychology*, *97*, 299-319. https://doi.org/10.1037/0022-0663.97.3.299

Pennington, B. F. (2009). *Diagnosing learning disorders: A neuropsychological framework. 2nd ed.* New York, NY: Guilford Press

Pennington, B. F., & Gilger, J. W. (1996). How is dyslexia transmitted? In C. H. Chase, G. D> Rosen & G. F. Sherman (Eds.), *Developmental dyslexia: Neural, cognitive, and genetic mechanisms* (pp. 41-61). Parkton, MD: York.

Peterson, R. L., & Pennington, B. F. (2010). Reading Disability. In O. K. Yeates, D. M. Ris, G.H. Taylor, & B. F. Pennington (Eds.), *Pediatric neuropsychology* (pp. 324-362). New York, NY: The Guliford Press.

Peterson, R. L., Pennington, B. F., & Olson, R. K. (2013). Subtypes of developmental dyslexia: testing the predictions of the dual-route and connectionist frameworks. *Cognition*, *126*, 20-38. https://doi.org/10.1016/j.cognition.2012.08.007

Pfost, M., Hattie, J., Dörfler, T., & Artelt, C. (2014). Individual differences in reading development: A review of 25 years of empirical research on Matthew effects in reading. *Review of Educational Research*, *84*, 203-244. https://doi.org/10.3102/0034654313509492

Phillips, B. M. & Lonigan, C. J. (2005). Social Correlates of Emergent Literacy. In M. J. Snowling and C. Hulme (Eds.). *The science of reading: A handbook* (pp. 173-187). Oxford: Blackwell Publishing.

Phillips, L. M., Norris, S. P., Osmond, W. C., & Maynard, A. M. (2002). Relative reading achievement: A longitudinal study of 187 children from first through sixth grades. *Journal of Educational Psychology*, *94*, 3-13. https://doi.org/10.1037/0022-0663.94.1.3

Pianta, R. C., & McCoy, S. J. (1997). The first day of school: The predictive validity of early school screening. *Journal of Applied Developmental Psychology*, *18*, 1-22. https://doi.org/10.1016/s0193-3973(97)90011-3

Planty, M., Hussar, W., Snyder, T., Provasnik, S., Kena, G., Dinkes, R., et al. (2008). The condition of education 2008 (NCES 2008-031). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://nces.ed.gov/pubs2007/2007065.pdf

Pollack, J.M., Atkins-Burnett, S., Najarian, M., and Rock, D.A. (2005a). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS–K), Psychometric Report for the Fifth Grade* (NCES 2006–036). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubs2006/2006036rev.pdf

Pollack, J., Atkins-Burnett, S., Rock, D., and Weiss, M. (2005b). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Psychometric Report for the Third Grade* (NCES 2005–062). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubs2005/2005062.pdf

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling. Quantitative applications in the social sciences.* Thousand Oaks, CA: Sage Publications, Inc.

Protopapas, A., Parrila, R., & Simos, P. G. (2016). In search of Matthew effects in reading. *Journal of Learning Disabilities*, *49*, 499-514. https://doi.org/10.1177/0022219414559974

Rathbun, A., and Zhang, A. (2016). *Primary Early Care and Education Arrangements and Achievement at Kindergarten Entry* (NCES 2016-070). National Center for Education Statistics, U.S. Department of Education. Washington, DC. Retrived from https://nces.ed.gov/pubs2016/2016070.pdf

Rathvon, N. (2004). *Early reading assessment: A practitioner's handbook.* New York: Guilford Press.

Rimm-Kaufman, S. E., Pianta, R. C., & Cox, M. J. (2000). Teachers' judgments of problems in the transition to kindergarten. *Early childhood Research Quarterly*, *15*, 147-166. https://doi.org/10.1016/s0885-2006(00)00049-1

Rock, D., Pollack, J., Atkins-Burnett, S., Hoffer, T., Meisels, S., Tourangeau, K., … Hausken, E. G. (2002). *Early Childhood Longitudinal Study- Kindergarten Class of 1998–99 (ECLS–K), Psychometric Report for Kindergarten Through First Grade* (NCES 2002-005). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubs2002/200205.pdf

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50*, 203-228. https://doi.org/10.1007/bf02294247

Rourke, B. P. (1975). Brain-behavior relationships in children with learning disabilities: A research program. *American Psychologist, 30,* 911-920. https://doi.org/10.1037/0003-066x.30.9.911

Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale, survey research tells us about student achievement: Insights from the prospects study of elementary school. *Teachers College Record, 104,* 1525-1567. https://doi.org/10.1037/e384482004-001

Rumberger, R. W., & Palardy, G. J. (2005). Does segregation still matter? The effect of student composition on academic achievement in high schools. *Teachers College Record,* 107, 1999-2045. https://doi.org/10.1111/j.1467-9620.2005.00583.x

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* New York: Wiley.

Rutter, M., Caspi, A., Fergusson, D., Horwood, L. J., Goodman, R., Maughan, B., ... & Carroll, J. (2004). Sex differences in developmental reading disability: New findings from 4 epidemiological studies. *Journal of American Medical Association*, *291*, 2007-2012. https://doi.org/10.1001/jama.291.16.2007

Rutter, M., & Yule, W. (1975). The concept of specific reading retardation. *Journal of Child*

Psychology and Psychiatry, *16*, 181-197. https://doi.org/10.1111/j.1469-7610.1975.tb01269.x

Sanford, A. K., Park, Y., & Baker, S. K. (2013). Reading growth of students with disabilities in the context of a large-scale statewide reading reform effort. *The Journal of Special Education*, *47*, 83-95. https://doi.org/10.1177/0022466911430350

Scarborough, H. S. (1990). Very early language deficits in dyslexic children. *Child Development*, *61*, 1728-1743. https://doi.org/10.2307/1130834

Scarborough, H. S. (1998). Early identification of children at risk for reading difficulties: Phonological awareness and some other promising predictors. In B.K. Shapiro, P. J. Accardo, & A. J. Capute (Eds.), *Specific reading disability: A review of the spectrum* (pp. 75-199). Timonium, MD: York Press.

Scarborough, H. S., & Parker, J. D. (2003). Matthew effects in children with learning disabilities: Development of reading, IQ, and psychosocial problems from grade 2 to grade 8. *Annals of Dyslexia*, *53*, 47-71. https://doi.org/10.1007/s11881-003-0004-6

Scammacca, N. K., Fall, A. M., & Roberts, G. (2015). Benchmarks for expected annual academic growth for students in the bottom quartile of the normative distribution. *Journal of Research on Educational Effectiveness*, *8*, 366-379. https://doi.org/10.1080/19345747.2014.952464

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* London: Chapman & Hall.

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research*, *8*, 3-15. https://doi.org/10.1177/096228029900800102

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147-177. https://doi.org/10.1037/1082-989x.7.2.147

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, *33*, 545-571. https://doi.org/10.1207/s15327906mbr3304_5

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, *96*, 265-282. https://doi.org/10.1037/0022-0663.96.2.265

Scheiber, C., Reynolds, M. R., Hajovsky, D. B., & Kaufman, A. S. (2015). Gender differences in achievement in a large, nationally representative sample of children and adolescents. *Psychology in the Schools*, *52*, 335-348. https://doi.org/10.1002/pits.21827

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523-568. https://doi.org/10.1037/0033-295x.96.4.523

Seltzer, M. H., Frank, K. A., & Bryk, A. S. (1994). The metric matters: The sensitivity of conclusions about growth in student achievement to choice of metric. *Educational Evaluation and Policy Analysis*, *16*, 41-49. https://doi.org/10.2307/1164382

Shankweiler, D., & Crain, S. (1986). Language mechanisms and reading disorder: A modular approach. *Cognition*, *24*, 139-168. https://doi.org/10.1016/0010-0277(86)90008-9

Share, D. L., Jorm, A. F., Maclean, R., & Matthews, R. (1984). Sources of individual differences in reading acquisition. *Journal of Educational Psychology, 76*, 1309-1324. https://doi.org/10.1037/0022-0663.76.6.1309

Share, D. L., & Stanovich, K. E. (1995). Cognitive processes in early reading development: Accommodating individual differences into a model of acquisition. *Issues in Education*, *1*, 1-57. https://doi.org/10.1016/j.lindif.2012.01.012

Shaywitz, S. (2003). *Overcoming dyslexia: A new and complete science-based program for reading problems at any level*. New York: Alfred A. Knopf.

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992a). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine*, *326*, 145-150. https://doi.org/10.1056/nejm199201163260301

Shaywitz, B. A., Fletcher, J. M., Holahan, J. M., & Shaywitz, S. E. (1992b). Discrepancy compared to low achievement definitions of reading disability: Results from the Connecticut longitudinal study. *Journal of Learning Disabilities*, *25*, 639-648. https://doi.org/10.1177/002221949202501003

Shaywitz, B. A., Holford, T. R., Holahan, J. M., Fletcher, J. M., Stuebing, K. K., Francis, D. J., & Shaywitz, S. E. (1995). A Matthew Effect for IQ but Not for Reading: Results from a Longitudinal Study. *Reading Research Quarterly*, *30*, 894-906. https://doi.org/10.2307/748203

Shaywitz, S. E., Morris, R., & Shaywitz, B. A. (2008a). The education of dyslexic children from childhood to young adulthood. *Annual Review of Psychology, 58*, 451-475. https://doi.org/10.1146/annurev.psych.59.103006.093633

Shaywitz, S. E., & Shaywitz, B. A. (2005). Dyslexia (specific reading disability). *Biological psychiatry*, *57*, 1301-1309. https://doi.org/10.1016/j.biopsych.2005.01.043

Shaywitz, S. E., & Shaywitz, B. A. (2008b). Paying attention to reading: The neurobiology of reading and dyslexia. *Development and Psychopathology, 20*, 1329-1349. https://doi.org/10.1017/s0954579408000631

Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M., & Escobar, M. D. (1990). Prevalence of reading disability in boys and girls: Results of the Connecticut Longitudinal Study. *Journal of American Medical Association, 264*, 998-1002. https://doi.org/10.1001/jama.264.8.998

Shaywitz, B., Shaywitz, S., Pugh, K., Mencl, W., Fulbright, R., Skudlarski, P., …, & Gore, J. (2002). Disruption of posterior brain systems for reading in children with developmental dyslexia. *Biological Psychiatry, 52*, 101-101. https://doi.org/10.1016/s0006-3223(02)01365-3

Shinn, M. R., Ysseldyke, J. E., Deno, S. L., & Tindal, G. A. (1986). A comparison of differences between students labeled learning disabled and low achieving on measures of classroom performance. *Journal of Learning Disabilities*, *19*, 545-552. https://doi.org/10.1177/002221948601900906

Sideridis, G.D., Morgan, P.L., Bostas, G., Padeliadu, S., & Fuchs, D. (2006). Predicting LD on the Basis of Motivation, Metacognition, and Psychopathology: A ROC analysis. *Journal of Learning Disabilities, 39*, 215-229. https://doi.org/10.1177/00222194060390030301

Siegel, L. S. (1992). An evaluation of the discrepancy definition of dyslexia. *Journal of Learning Disabilities*, *25*, 618–629. https://doi.org/10.1177/002221949202501001

Siegel, L. S., & Himel, N. (1998). Socioeconomic status, age and the classification of dyslexics and poor readers: The dangers of using IQ scores in the definition of reading disability. *Dyslexia*, *4*, 90-104. https://doi.org/10.1002/(sici)1099-0909(199806)4:2<90::aid-dys105>3.0.co;2-3

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*, 417-453. https://doi.org/10.3102/00346543075003417

Skibbe, L. E., Grimm, K. J., Stanton-Chapman, T. L., Justice, L. M., Pence, K. L., & Bowles, R. P. (2008). Reading trajectories of children with language difficulties from preschool through fifth grade. *Language, Speech, and Hearing Services in Schools*, *39*, 475-486. https://doi.org/10.1044/0161-1461(2008/07-0016)

Smith, J. R., Brooks-Gunn, J., & Klebanov, P. (1997). Consequences of living in poverty for young children's cognitive and verbal ability and early school achievement. In G. J. Duncan & J. Brooks-Gunn (Eds.), *Consequences of growing up poor* (pp. 132-189). New York: Russell Sage Foundation.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Snowling, M. J. (1981). Phonemic deficits in developmental dyslexia. *Psychological Research*, *43*, 219-234. https://doi.org/10.1007/bf00309831

Snowling, M. J., Gallagher, A., & Frith, U. (2003). Family risk of dyslexia is continuous: Individual differences in the precursors of reading skill. *Child Development*, *74*, 358-373. https://doi.org/10.1111/1467-8624.7402003

Spear-Swerling, L., & Sternberg, R. J. (1996). *Off track: When poor readers become "learning disabled*." Boulder, CO: Westview Press.

Specht, K., Hugdahl, K., Ofte, S., NygÅrd, M., BjØrnerud, A., Plante, E., & Helland, T. (2009). Brain activation on pre-reading tasks reveals at-risk status for dyslexia in 6-year-old children. *Scandinavian Journal of Psychology*, *50*, 79-91. https://doi.org/10.1111/j.1467-9450.2008.00688.x

Stanovich, K. E. (1985). Explaining the variance in reading ability in terms of psychological processes: What have we learned? *Annals of Dyslexia*, *35*, 67-96. https://doi.org/10.1007/bf02659181

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21,* 360-407. Retrieved from http://psycnet.apa.org/record/1987-15181-001

Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities*, *21*, 590-604. https://doi.org/10.1177/002221948802101003

Stanovich, K. E. (1991). Discrepancy Definitions of Reading Disability: Has Intelligence Led Us Astray? *Reading Research Quarterly*, *26*, 7-29. https://doi.org/10.2307/747729

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology*, *86*, 24-53. https://doi.org/10.1037/0022-0663.86.1.24

Stein, J., & Walsh, V. (1997). To see but not to read; the magnocellular theory of dyslexia. *Trends in Neurosciences*, *20*, 147-152. https://doi.org/10.1016/s0166-2236(96)01005-3

Stevenson, J. (1988). Which aspects of reading ability show a 'hump' in their distribution?. *Applied Cognitive Psychology*, *2*, 77-85. https://doi.org/10.1002/acp.2350020107

Stipek, D., & Byler, P. (2001). Academic achievement and social behaviors associated with age of entry into kindergarten. *Journal of Applied Developmental Psychology*, *22*, 175-189. https://doi.org/10.1016/s0193-3973(01)00075-2

Stuebing, K. K., Barth, A. E., Trahan, L. H., Reddy, R. R., Miciak, J., & Fletcher, J. M. (2015).

Are child cognitive characteristics strong predictors of responses to intervention? A meta-analysis. *Review of Educational Research*, *85*, 395-429. https://doi.org/10.3102/0034654314555996

Stuebing, K. K., Fletcher, J. M., LeDoux, J. M., Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2002). Validity of IQ-discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal*, *39*, 469-518. https://doi.org/10.3102/00028312039002469

Svetaz, M. V., Ireland, M., & Blum, R. (2000). Adolescents with learning disabilities: Risk and protective factors associated with emotional well-being: Findings from the National Longitudinal Study of Adolescent Health. *Journal of Adolescent Health*, *27*, 340-348. https://doi.org/10.1016/s1054-139x(00)00170-1

Swank, L. K., & Catts, H. W. (1994). Phonological awareness and written word decoding. *Language, Speech, and Hearing Services in Schools*, *25*, 9-14. https://doi.org/10.1044/0161-1461.2501.09

Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain and language*, *9*, 182-198. https://doi.org/10.1016/0093-934x(80)90139-x

Tallal, P., Miller, S., & Fitch, R. H. (1995). Neurobiological basis of speech: A case for the preeminence of temporal processing. *The Irish Journal of Psychology*, *16*, 194-219. https://doi.org/10.1111/j.1749-6632.1993.tb22957.x

Thabane, L., Mbuagbaw, L., Zhang, S., Samaan, Z, Marcucci, M., Ye, C., … Goldsmith, C. (2013). A tutorial on sensitivity analyses in clinical trials: The what, why, when and how. *BMC Medical Research Methodology, 13,* 1-12. https://doi.org/10.1186/1471-2288-13-92

Tiu, R. D., Thompson, L. A., & Lewis, B. A. (2003). The role of IQ in a component model of reading. *Journal of Learning Disabilities*, *36*, 424-436. https://doi.org/10.1177/00222194030360050401

Tomalski, P., Moore, D. G., Ribeiro, H., Axelsson, E. L., Murphy, E., Karmiloff-Smith, A., … Kushnerenko, E. (2013). Socioeconomic status and functional brain development - associations in early infancy. *Developmental Science*, *16*, 676-687. https://doi.org/10.1111/desc.12079

Tong, X., Deacon, S. H., Kirby, J. R., Cain, K., & Parrila, R. (2011). Morphological awareness: A key to understanding poor reading comprehension in English. *Journal of Educational Psychology*, *103*, 523-534. https://doi.org/10.1037/a0023495

Torgesen, J. K. (1999). Phonologically based reading disabilities: Toward a coherent theory of one kind of learning disability. In R. J. Sternberg & L. Spear-Swerling (Eds.), *Perspectives on learning disabilities* (pp. 231–262). New Haven: Westview.

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice*, *15*, 55-64. https://doi.org/10.1207/sldrp1501_6

Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology*, *40*, 7-26. https://doi.org/10.1016/s0022-4405(01)00092-9

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. K., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities*, *34*, 33-58. https://doi.org/10.1177/002221940103400104

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in second-to fifth-grade children. *Scientific Studies of Reading*, *1*, 161-185. https://doi.org/10.1207/s1532799xssr0102_4

Tourangeau, K., Brick, M., Le, T., Wan, S., Weant, M., Nord, C., … Walston, J. (2004). *User's manual for the ECLS-K third grade, public-use data file and electronic code book (NCES 2004-001).* National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved from https://nces.ed.gov/pubs2004/2004001.pdf

Tourangeau, K., Burke, J., Le, T., Wan, S., Weant, M., Nord, C., … Walston, J. (2002). *User's manual for the ECLS-K first grade, public-use data file and electronic code book (NCES 2002-135).* National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved from https://nces.ed.gov/pubs2002/2002135_1.pdf

Tourangeau, K., Nord, C., Le, T., Pollack, J.M., & Atkins-Burnett, S. (2006). *Early childhood longitudinal study, kindergarten class of 1998–99 (ECLS-K), combined user's manual for the ECLS-K fifth grade data files and electronic codebooks* (NCES 2006–032). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubs2006/2006032.pdf

Tourangeau, K., Nord, C., Le, T., Sorongon, A. G., & Najarian, M. (2009). *Early childhood longitudinal study, kindergarten class of 1998–99 (ECLS-K), combined user's manual for the ECLS-K eighth grade and K–8 full sample data files and electronic codebooks* (NCES 2009–004). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved from https://nces.ed.gov/ecls/data/ECLSK_K8_Manual_part1.pdf

UCLA: Statistical Consulting Group (2017). FAQ: How are the likelihood ratio, Wald, and Lagrange multiplier (score) tests different and/or similar?. Retrieved from https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faqhow-are-the-likelihood-ratio-wald-and-lagrange-multiplier-score-tests-different-andor-similar/

Unnever, J. D., Kerckhoff, A. C., & Robinson, T. J. (2000). District variations in educational resources and student outcomes. *Economics of Education Review*, *19*, 245-259. https://doi.org/10.1016/s0272-7757(99)00043-6

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP). (2016). NAEP Data Explorer [Data file]. Retrieved from https://nces.ed.gov/nationsreportcard/naepdata/

U.S. Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs. (2015). *37th Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act, 2015,* Washington, D.C. Retrieved from https://www2.ed.gov/about/reports/annual/osep/2015/parts-b-c/37th-arc-for-idea.pdf

Van der Wissel, A., & Zegers, F. E. (1985). Reading retardation revisited. *British Journal of Developmental Psychology*, *3*, 3-9. https://doi.org/10.1111/j.2044-835x.1985.tb00947.x

Van Steensel, R. (2006). Relations between socio-cultural factors, the home literacy environment and children's literacy development in the first years of primary education. *Journal of Research in Reading*, *29*, 367-382. https://doi.org/10.1111/j.1467-9817.2006.00301.x

Vaughn, S., Levy, S., Coleman, M., & Bos, C. S. (2002). Reading instruction for students with LD and EBD: A synthesis of observation studies. *The Journal of Special Education*, *36*, 2-13. https://doi.org/10.1177/00224669020360010101

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional children*, *69*, 391-409. https://doi.org/10.1177/001440290306900401

Vaughn, S., Moody, S. W., & Schumm, J. S. (1998). Broken promises: Reading instruction in the resource room. *Exceptional children*, *64*, 211-225. https://doi.org/10.1177/001440299806400205

Vaughn, S., Wexler, J., Leroux, A., Roberts, G., Denton, C., Barth, A., & Fletcher, J. (2012). Effects of intensive reading intervention for eighth-grade students with persistently inadequate response to intervention. *Journal of Learning Disabilities*, *45*, 515–525. https://doi.org/10.1177/0022219411402692

Vellutino, F. (1979). *Dyslexia: Theory and research.* Cambridge, MA: MIT Press

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): what have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*, 2-40. https://doi.org/10.1046/j.0021-9630.2003.00305.x

Vellutino, F. R., Scanlon, D. M., & Lyon, G. R. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the IQ-achievement discrepancy definition of reading disability. *Journal of Learning Disabilities*, *33*, 223-238. https://doi.org/10.1177/002221940003300302

Vellutino, F., Scanlon, D. M., & Lyon, G. R. (2003). Differentiating between difficult-to-remediate poor readers. *Journal of Learning Disabilities*, *33*, 223-239. https://doi.org/10.1177/002221940003300302

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology*, *88*, 601-638. https://doi.org/10.1037//0022-0663.88.4.601

Vellutino, F. R., Scanlon, D. M., Small, S. & Fanuele, D. P. (2006). Response to intervention as a vehicle for distinguishing between children with and without reading disabilities: Evidence for the role of kindergarten and first-grade interventions. *Journal of Learning Disabilities, 39*, 157-169. https://doi.org/10.1177/00222194060390020401

Vellutino, F. R., Scanlon, D. M., & Spearing, D. (1995). Semantic and phonological coding in poor and normal readers. *Journal of Experimental Child Psychology*, *59*, 76-123. https://doi.org/10.1006/jecp.1995.1004

Vellutino, F. R., Scanlon, D. M., Zhang, H., & Schatschneider, C. (2008). Using response to kindergarten and first grade intervention to identify children at-risk for long-term reading difficulties. *Reading and Writing*, *21*, 437-480. https://doi.org/10.1007/s11145-007-9098-2

Von Hippel, P. T. (2012). Should a normal Imputation model be Modified to impute skewed variables? *Sociological Methods & Research*, *42*, 105-138. https://doi.org/10.1177/0049124112464866

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological bulletin*, *101*, 192-212. https://doi.org/10.1037/0033-2909.101.2.192

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology*, *85*, 83-103. https://doi.org/10.1037/0022-0663.85.1.83

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental psychology*, *30*, 73-87. https://doi.org/10.1037/0012-1649.30.1.73

Walberg, H. J., & Tsai, S. L. (1983). "Matthew" effects in education. *American Educational*

Research Journal, *20*, 359-373. https://doi.org/10.2307/1162605

Wenglinsky, H. (1998). Finance equalization and within-school equity: The relationship between education spending and the social distribution of achievement. *Educational Evaluation and Policy Analysis*, *20*, 269-283. https://doi.org/10.2307/1164325

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child development*, *69*, 848-872. https://doi.org/10.2307/1132208

Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods*, *8*, 16-37. https://doi.org/10.1037/1082-989x.8.1.16

Williamson, G. L., Appelbaum, M., & Epanchin, A. (1991). Longitudinal analyses of academic achievement. *Journal of Educational Measurement*, *28*, 61-76. https://doi.org/10.1111/j.1745-3984.1991.tb00344.x

Wolf, M., & Bowers, P.G. (1999). The double deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, *91,* 1-24. https://doi.org/10.1037/0022-0663.91.3.415

Wolf, M., Bowers, P.G., & Biddle, K. (2000). Naming-speed processes, timing, and reading: a conceptual review. *Journal of Learning Disabilities, 33*, 387-407. https://doi.org/10.1177/002221940003300409

Wolf, M., & Gow, D. (1986). A longitudinal investigation of gender differences in language and reading development. *First Language*, *6*, 81-110. https://doi.org/10.1177/014272378600601701

Wu, Q., Morgan, P. L., & Farkas, G. (2014). Does Minority Status Increase the Effect of Disability Status on Elementary Schoolchildren's Academic Achievement?. *Remedial and Special Education*, *35*, 366-377. https://doi.org/10.1177/0741932514547644

Yoshikawa, H., Aber, J. L., & Beardslee, W. R. (2012). The effects of poverty on the mental, emotional, and behavioral health of children and youth: Implications for prevention. *American Psychologist*, *67*, 272-284. https://doi.org/10.1037/a0028015

Ysseldyke, J. E., Algozzine, B., Shinn, M. R., & McGue, M. (1982). Similarities and differences between low achievers and students classified learning disabled. *The Journal of Special Education*, *16*, 73-85. https://doi.org/10.1177/002246698201600108

Zehler, A. M., Fleischman, H. L., Hopstock, P. J., Stephenson, T. G., Pendzick, M. J., & Sapru, S. (2003). Descriptive study of services to LEP students and LEP students with disabilities: Research Report Arlington, VA: Development Associates, Inc. Retrieved from http://ncela.us/files/rcd/BE021199/special_ed4.pdf