2012-07-16

Evaluation of the Effectiveness of the Students and Teachers Achieving Reading Success Program for First Graders

Whitney Ann Phillips
Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Educational Assessment, Evaluation, and Research Commons

Evaluation of the Effectiveness of the Students and Teachers Achieving

Reading Success Program for First Graders

Whitney Ann Phillips

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Richard R Sudweeks, Chair
Janet Young
Timothy Morrison
Lane Fischer
Joseph Olsen

Educational, Inquiry, Measurement, and Evaluation

Brigham Young University

August 2012

ABSTRACT

Evaluation of the Effectiveness of the Students and Teachers Achieving
Reading Success Program for First Graders

Whitney Ann Phillips
Educational Inquiry, Measurement, and Evaluation, BYU
Doctor of Philosophy

Most students progress in learning when school is in session. However, during the summer months formal education often ends, and many of the gains students make during the academic year are lost over the summer break. The Alpine School District developed the *Students and Teachers Achieving Reading Success* (STARS) program, an extensive summer reading program for struggling readers. The purpose of this study was to evaluate the effectiveness of the STARS program on reading ability for students exiting the first grade, as measured by the *Developmental Reading Assessment 2* (DRA2). Results from a mixed-method ANOVA indicated that STARS students performed better than a nonequivalent control group ($p < .001$). Results from the multilevel growth modeling analysis provide evidence that the STARS participants performed better than those who were eligible for the program but did not participate. STARS participants improved in their reading ability at a significantly higher rate than students who were not eligible for the program and did not participate. The results indicated that the reading achievement gap of STARS participants narrowed by the end of second grade. Moderation variables were not statistically significant in their impact of reading trajectories between STARS participants and nonparticipants.

ACKNOWLEDGEMENTS

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Since 2001, the federal No Child Left Behind Act (NCLB) has held states accountable for eliminating achievement gaps by ensuring that all students are proficient in reading and mathematics. Duncan et al., (2006) refer to the achievement gap as the following:

> the observed disparity on a number of educational measures between the performance of groups of students, especially groups defined by gender, race/ethnicity, and socioeconomic status. The achievement gap can be observed on a variety of measures, including standardized test scores, grade point average, dropout rates, and college enrollment and completion rates. (p. 1429)

NCLB requires all districts receiving federal funds to document adequate yearly progress (AYP) in reading for all students or risk having federal funds withheld, and parents may choose to have their children attend schools where AYP has been made (No Child Left Behind Act of 2001, 2002). In 2004, Congress provided educators one way to close the achievement gap: the Response to Intervention (RTI) initiative. This initiative has two purposes: (a) to provide increasingly intensive expert reading instruction to ensure that students having difficulty learning to read are not simply getting too little or too inferior reading instruction, and (b) to locate students who exhibit difficulties even after receiving intensive reading instruction (Allington, 2011).

All 50 states currently have developed standards for K-12 education and have implemented mandatory assessments in grades 3 and higher to determine whether or not these standards have been met. While few states have mandated assessment at the primary level, many have adopted K-2 literacy initiatives that strongly recommend some form of assessment in these grades as well (Kagan & Scott-Little, 2004). As part of NCLB legislation, the federal

government offered substantial grants to states willing to implement its Reading First initiative, which focuses on early identification and remediation of students at risk of not meeting reading standards when accountability testing begins. States who received these federal grants were required to administer screening and diagnostic assessments to determine which students in grades K-3 were at risk of reading failure (NCLB, 2002). However, Reading First was not as successful as it was hoped to be. In fact, Gamse, Bloom, Kemple, and Jacob (2008) found that children in Reading First did no better than children in comparison groups.

In addition to federal legislation, individual school districts have created a variety of ways to help their students' loose less reading ability during the summer months. With the added pressure of recent budget cuts, school districts are interested in funding summer reading programs that have proven to be effective. Therefore, it is essential that school districts monitor student progress in the primary grades and evaluate their summer reading programs (McAfee & Leong, 2002).

Problem Statement

The Alpine School District (ASD) in Utah County, Utah, has developed *Students and Teachers Achieving Reading Success* (STARS), an intensive summer reading program for struggling readers who have just completed kindergarten, first, or second grade (K-2). The ASD first implemented STARS in 2000, and it has served approximately 2,500 children. The National Reading Panel (2001) suggests that teachers are expected to provide best-practice instruction including explicit, intense, and systematic instruction in phonemic awareness, phonics, fluency, vocabulary, and comprehension. Teachers and program administrators believe STARS has had significant positive effects on student reading levels because of anecdotal evidence. However, in an age of increased budgetary challenges for public education, the future of STARS depends

largely on the ability of school districts to demonstrate the efficacy of the program and the ultimate advantage to the school district.

Research Questions

The purpose of this research study was to use statistical tools to analyze existing data on reading level performance to examine the effect of the STARS program on participants who had just finished the first grade. Specifically, this study focused on three research questions:

1. What effect does participation in the STARS program have on students' reading ability as measured by the Developmental Reading Assessment 2, between the first-grade spring administration and the second-grade fall administration?

2. How does participation in the STARS program affect students' reading growth as measured by differences in their first grade reading growth trajectory and their second grade reading growth trajectory?

3. To what extent is the relationship between first grade students' participation or nonparticipation in the STARS program and their post-treatment reading test score trajectories moderated by students' gender, socioeconomic status, or ethnicity?

Chapter 2: Literature Review

Reading is critical for academic achievement and career success. Lyon (1998) remarks on the importance in reading:

The child and adult who cannot read at a comfortable level experience significant difficulty mastering many types of academic content, are at a substantial risk for failure in school, and are frequently unable to reach their potential in the vocational and occupational arena. (p. 1)

Children who have not been identified as having special needs are often not given adequate intervention to improve their reading abilities. Without intervention, struggling readers fall further and further behind their classmates. In view of this problem, this review of literature will (a) describe struggling readers, (b) summarize predictors of struggling readers, (c) provide evidence of summer reading loss, (d) review research on effective reading instruction for struggling readers, (e) evaluate summer reading programs, (f) discuss and analyze reading assessments, and (g) suggest multilevel modeling as an appropriate tool to evaluate program effectiveness.

Factors Related to Struggling Readers

Public schools in the United States are open to all students regardless of their gender, race/ethnicity, or socioeconomic background. NCLB requires these schools to provide equal access to a quality education to all students by removing barriers to learning, especially among students from disadvantaged backgrounds (NCLB, 2002). Under NCLB, the federal government requires all states receiving federal funds to put into place a set of standards, together with a detailed testing plan, to ensure that all children meet these standards. With its promise of holding schools accountable for educating all children, federal legislators hoped to level the playing field

by increasing funding for schools that serve poor children, by ensuring that all children receive instruction from highly qualified teachers, and by monitoring accountability through disaggregation of achievement data. Schools that fail to meet AYP for all children are subject to sanctions (Oakes, 2005).

The rules that were designed to hold schools accountable for all children regardless of gender, race, English proficiency, socioeconomic status, or disability are criticized for being counterproductive. According to Kane and Staiger (3002) "subgroup targets cause large numbers of schools to fail, arbitrarily single out schools with large minority subgroups for sanctions and exclude them from awards, or statistically disadvantage diverse schools that are more likely to be attended by minority students" (p. 174). Many people hoped that the focus on the achievement gap would result in improvements in schools for our lowest-performing students.

For some children, learning to read is very difficult and unrewarding (Allington, 2011). Learners who struggle with reading often have deficits associated with (a) decoding and analyzing word structures, (b) comprehending written texts, and (c) analyzing and reading words (Catts & Hogan, 2003). McGee and Richgels (2008) believe that "some children seem to struggle to acquire literacy even within literacy-rich classrooms and with a wide variety of instructional experiences" (p. 320). This study focuses on three variables that have been found to be related to struggling readers: gender, ethnicity, and socioeconomic status.

Gender. Research on the differences in reading achievement between boys and girls has produced conflicting results with many questions remaining unresolved. Most studies recognize that gender differences remain with gaps in reading favoring girls over boys and a larger number of males suffering from language disabilities such as dyslexia (Ely, 2005). In 2005 National Assessment of Educational Progress indicated a slightly higher reading score for girls than boys

in fourth grade, with 34% of girls scoring proficient or above as compared to 29% of boys. These gaps have remained unchanged for every reported year since 1992 (U.S. Department of Education, 2005).

Ready, LoGerfo, Burkham, and Lee (2005) analyzed longitudinal data including 16,883 kindergartners (8,701 boys and 8,182 girls). The assessment instrument *Early Childhood Longitudinal Study-Kindergarten* was individually administered to assess literacy ability. Results indicated that girls entered kindergarten with stronger literacy skills than boys and that they learned slightly more during kindergarten (i.e., girls gained an average of 10.3 points while boys gained an average of 9.6 points). A significant portion of the gender difference in literacy skills at the end of kindergarten was attributed to skill differences present when students entered kindergarten. About 70% of the gap between boys and girls was attributed to girls' learning approaches (e.g., attentiveness, task persistence, eagerness to learn, learning independence, flexibility, and organization). Boys were more likely to be recommended for retention in kindergarten than girls (i.e., 5.7% of boys vs. 3.3% of girls).

McNiece, Bidgood, and Soan (2004) analyzed trends in reading achievement based on gender in two longitudinal studies in 1974 and 1986 from Great Britain. The results of these two studies were contradictory. Using data from a national child development study, the researchers found that girls were better readers than boys in the early primary grades, but boys had caught up by the end of primary education and had slightly surpassed girls by age 16. However, data from the British Cohort Study indicated that boys were better readers than girls in first grade but that girls had surpassed boys by the end of primary and maintained that through age.

Socioeconomic status. Educators have long known that there is a relationship between socioeconomic status (SES) and achievement. Only 16% of students qualified for free or reduced

lunch compared with 42% of students not eligible for this program scored in the proficient range in reading (NAEP, 2005). This achievement gap has changed little since 1998. Studies consistently find that students who are subjected to long-term poverty or who attend schools in which a high percentage of students are poor have lower achievement test scores (Nyhan & Alkadry, 1999; Oakes, 2005; White, Reynolds, Thomas, & Gitzlaff, 1993). According to Sirin (2005) family socioeconomic status is the strongest indicator of academic achievement.

Linnakylä, Malin, and Taube (2004) report that low reading literacy achievement is significantly correlated with male gender, immigrant status, low SES background, several siblings, low academic self-esteem, pressure to achieve, strong effort and perseverance as well as lack of engagement in reading and a frequent use of computers. Linnakylä, Malin, and Taube (2004) stated "Hence the economic, cultural and social capital of the family does influence the children's learning in various ways, either promoting or hindering it" (p. 233).

Poverty is among the leading factors that predict poor reading achievement (Aikens, Barbarin, 2008; Snow, Burns, & Griffin, 1998). SES affects families, schools, and neighborhoods. These environments often include poor literacy resources, limited displayed print, and poor public libraries (Neuman, Celano, Greco, & Shue, 2001). A longitudinal study of 368 elementary students in Baltimore indicated that lower SES students who were academically behind their upper SES peers in first grade were even further behind five years later (Entwisle, Alexander, & Olson, 2005).

Several studies have demonstrated that the achievement gap between high and low SES students widens more over the summer, when children are not in school, than it does during the school year (Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996; Entwisle et al., 2005). During the summer months, when school is not in session, children with low SES lose about two school

months' worth of reading and math skills. Therefore, when these children return to school, it takes an average of two months to get the students back to the academic level that they reached the previous academic year (The Progress of Education Reform, 2009). In conclusion, SES does appear to impact students' academic achievement, and much of the widening of the achievement gap between higher and lower SES students appears to occur during non-instructional times.

McCoach, O'Connell, Reis, and Levitt (2006) found that SES had a minimal impact on reading growth while school was in session; however, it had a larger impact on summer reading growth. These results suggest the potential importance of preschool and summer programs for children from lower SES backgrounds. McCoach et al. (2006) used hierarchical linear modeling to map the trajectory of children's reading development over the kindergarten and first grade years. Based on the knowledge that children begin kindergarten at different levels of reading ability, the study was designed to illuminate individual growth patterns for the purpose of understanding and addressing the persistent gaps in reading achievement. The data were drawn from the first four waves of the Early Childhood Longitudinal Study-Kindergarten cohort (ECLS-K), which enabled researchers to examine the variables: SES, race and ethnicity, gender, and age at kindergarten entry, and school level effects and academic achievement.

Their analysis was based on a 3-level growth curve model examining time, student, and school effects. The findings confirmed significant individual differences exist in reading ability when children begin kindergarten. On average, children from higher income backgrounds began school with higher reading performance. Conversely, schools with children at significantly higher levels of poverty had lower initial reading performance.

Children with higher reading scores made more progress during the summer months. These children also tended to be from higher income families. As interpreted by McCoach et al.

(2006), "the widening of the gap between good and poor schools may be occurring during the summer months" (p. 25). The achievement gap between poorer and more affluent students existed when children started school, expanded during kindergarten, and most significantly, grew even more over the summer. While reading growth over the summer was small on average, more affluent students made more progress. In fact, in all analyses, SES emerged as one of the foremost factors in the children's reading ability when they began kindergarten and the progress they made over the summer months. McCoach et al. (2006) joined the many researchers who advocate the expansion of summer learning experiences for low-income students. In order to mitigate the achievement gap in reading, McCoach et al. (2006) recommended literacy programs for low-income preschool age children as well as summer enrichment activities.

Ethnicity. In 2010, minorities made up 42 percent of the U.S. population (Bates, 2011). Hispanics were the largest minority group, representing 18 percent of the population followed by Blacks at 13 percent. According to the 2005 NAEP results, minority students are still at a disadvantage in U.S. schools. Differences in the experiences of minority children prior to entering kindergarten contribute to reading achievement differences in the primary grades. For example, prior to beginning kindergarten, white children are much more likely than African American or Hispanic children to be read to in the home or to be taken on trips to the library (Hoffman & Liagas, 2003). Children with different opportunities to learn and practice early reading skills may contribute to achievement differences across ethnic categories. For example, fewer than 7% of White mothers of school-age children have less than a high school degree, while almost 20% of African American mothers and almost 50% of Hispanic mothers have not completed high school (National Center for Education Statistics, 2002).

In 2001, the *Developmental Reading Assessment* 2 was administered to 2,564 second graders from Durham, North Carolina. Whites performed better than all other ethnic groups. Hispanic students scored the lowest on the text reading level but had the highest rate of growth, followed by African Americans (Durham Public School Office of Research and Accountability, 2001).

Longitudinal trends in reading achievement according to race/ethnicity were examined in a study by McNiece, Bidgood, and Sloan (2004). Consistent with other studies, findings by these researchers were that ethnic minority groups do not perform as well as whites in reading trends. However, in the British Cohort Study (1970-1986), the ethnic minority group made significantly more progress than the white group over the same time period (13.3% gain in percent difference in mean scores).

On the 2003 California Stanford 9 student achievement test in the Pasadena Unified School District, there was a 27% difference between the achievement of white and black students (Bali & Alvarez, 2004). The study included 1147 fourth-grade students from the Pasadena Unified School District, a large, racially diverse system in California. Bali and Alvarez found that in the first grade, the average reading score of Black students was over six points lower than that of White students. Hispanic students scored 13 points lower than White students. By fourth grade, the gap in Hispanic/White reading scores had decreased slightly while the Black/White gap had increased. The Black/White gap was statistically significant throughout and increased each year. By fourth grade, the gap between Blacks and Whites in reading was twice as wide as the gap between Hispanics and Whites. The authors suggest that the gaps for Hispanics are caused by school factors and the interaction of language and school factors, while the

Black/White gap suggests that family and preschool factors play a stronger role (Bali & Alvarez, 2004).

McCoach, et al. (2006) found that minority students began Kindergarten with scores just slightly lower than White students. However, during the summer months, minority students lost significantly more reading ability than did their White classmates. These findings support the claim made by Cooper, Chalton, Valentine, Mudlenbruck, and Borman (2000) who suggest that children at risk for academic failure lose more reading ability over the summer months than their classmates who are not at risk.

Summer Reading Loss

Summer reading loss, also known as *summer set back* or *summer slide*, refers to the lack of achievement gain or even a decrease of achievement over the summer months when school is not in session (The Progress of Education Reform, 2009). Available research indicates that summer reading loss for poorer children is greater than for children from wealthier families. On average, students who have a low SES experience two months of reading loss each summer. In contrast, students from middle- to higher-income families experience a gain in reading skills over the summer (Cooper, et al., 2000). Therefore, at-risk children fall further and further behind in the reading levels of their peers during the summer months, widening the gap between students who are academically successful and those who are not.

There are a variety of reasons why summer reading loss is more prominent among poorer children. Poorer children have less access to books when they are not in school. These children get most of their books from school, and when school is not in session, they do not have access to books. School libraries have fewer open hours and more restrictions on checking out books in poorer neighborhood schools. Wealthier communities have more books and bookstores than in

poorer neighborhoods. The fact that there are fewer bookstores in poorer neighborhoods may be related to the fact that poorer families with less discretionary money will most likely purchase fewer books (Neuman & Celano, 2001).

Book access may not be the sole reason that poorer children have greater summer reading loss. Allington and McGill-Franzen (2003) suggest that motivation may influence a child's choice in voluntary reading over the summer. For example, children's self-efficacy in reading may influence if they are motivated to read. Children struggling with reading are less likely to choose voluntarily to read during the summer when reading is not required by a teacher.

Effective Reading Instruction for Struggling Readers

The Progress of Education Reform (2009) states that "without ongoing opportunities to learn and practice essential skills, kids fall behind on measures of academic achievement over the summer months" (p. 1). Reading loss is most severe for low-income youth. Parents, teachers, school administrators, and researchers have put forth an increasing amount of effort and resources into creating effective summer school programs to alleviate summer reading loss of students at risk. Students who struggle with reading skills in early elementary school will most likely experience difficulty with reading throughout their adulthood (Allington, 2011, Gerston, 1996; Lyon, 1998). Reading problems persist if students do not receive appropriate, effective instruction. This section will review the theoretical propositions and instructional guidelines of teaching reading. It also includes information on early intervention strategies, explicit and systematic instructional strategies, and instructional implications for teaching reading to struggling readers.

Theoretical propositions. Cambourne (2002) lists three theoretical propositions for effective reading instruction. First, what is learned cannot be separated from the context in which

12

it is learned. This means that the "experiences and contents in which learning to read is embedded will be critical to each learner's understanding of, and ability to use, reading" (Cambourne, 2002, p. 27).

Second, the purposes or goals that the learner brings to the learning situation are central to potential learning opportunities. A teacher who considers a child's purpose in learning has the perspective that a child brings value to a learning experience, instead of the behaviorist point of view that a child is a clean slate that must be trained. A child makes a decision whether to be engaged in reading instruction or not. Engagement "incorporates a range of different behaviors. It has overtones of attention; learning is unlikely if learners do not attend to demonstrations in which they are immersed" (Cambourne, 2002, p. 27-28). By creating a need or purpose in reading, a child is more likely to be attentive and engaged in learning. Cambourne (2002) discussed the following principles of engagement: (a) learners are most likely to engage deeply with demonstrations if they believe that they are capable of ultimately learning or doing whatever is being demonstrated, (b) learners are most likely to engage deeply with demonstrations if they believe that learning whatever is being demonstrated has some potential value, purpose, and use for them, (c) learners are more likely to engage with demonstrations if they are free from anxiety, and (d) learners are more likely to engage with demonstrations given by someone they like, respect, admire, trust, and would like to emulate. These principles mean that teachers must know how to create learning environments that encourage learners to engage as deeply as possible.

Third, knowledge and meaning are socially constructed through the processes of negotiation, evaluation, and transformation. One implication of social construction is that teachers should use collaborative groups because these groups provide a means by which

13

children can learn by the modeling of other children. Children are also able to assess their own performance compared to the performance of their peers.

Instructional guidelines for teaching reading. The Partnership for Reading (2003) that includes the National Institute for Literacy, the National Institute of Child Health and Human Development, the U.S. Department of Education, and the U.S. Department of Health and Human Services, recommended the following guidelines for the teaching of reading:

1. A comprehensive reading program is grounded on scientifically based research, with all the components of the program carefully aligned so that instruction is seamlessly organized.

2. Instructional materials are geared to the specific needs of the children in that school. Administrators should ensure the use of the materials that provide highly explicit and systematic instruction.

3. Highly qualified initial training and ongoing staff development is provided for teachers that focuses on the foundational concepts of learning to read and the use of a selected comprehensive reading program. Professional development must focus on helping the teacher apply the proven principles of effective classroom reading instruction.

4. Adequate and uninterrupted time must be provided for reading instruction. Too often, schools allocate a sufficient quantity of time but allow it to be broken up, which is not effective. Also, children who are behind must be provided extra instructional time.

5. A system for regular assessment of student progress should be in place throughout the school year, using valid and reliable classroom-based instructional assessments to determine whether goals are being reached by the expected time.

6.   Data should be used from classroom assessments to determine where help is needed at a classroom, school, or district level. All educators must be trained to use data to make appropriate and effective instructional decisions.

7.   Intervention must be provided when student progress is not adequate, rather than when it is not at desired levels. The intervention should provide help that aligns with the overall reading program and targets the identified areas where the student requires additional instruction.

Early intervention strategies. Early intervention has been hailed as a preventative measure for struggling readers. Early intervention typically occurs in small groups. The assumption is that smaller groups present greater opportunities for students to socializing with their peers and learning from them. A teacher often scaffolds learning for the children based on their individual needs. Through dialogues and shared experiences, children acquire the cognitive processes necessary for future learning (Bodrova & Leong, 1996).

Strickland et al. (2002) recommend four components of early intervention programs from grades 1-3:

1.   Reading comprehension strategies focus on the self-monitoring of texts and the use of fix-up strategies when misunderstandings occur. Some of the ways that children demonstrate their comprehension of texts involve retelling stories, making predictions, summarizing stories and books, and participating in discussions generated by comprehension questions.

2.   Word recognition strategies, including phonics and structural analysis, are addressed as essentials for skillful reading. Most programs stress phonics as a tool to decode words.

3.   Fluency is addressed by helping students develop the ability to read expressively and meaningfully, as well as accurately, with appropriate speed.

4.   Writing is taught as an important path to word analysis skills, spelling, and self-expression. Writing is seen as a key to understanding the relationship between oral language and print. (p. 77)

Explicit and systematic instructional strategies. Explicit instruction has also been helpful in teaching struggling readers. McGee and Richgels (2008) define explicit instruction as "teaching that includes specifying of learning outcomes, modeling of processes, thinking aloud, and explaining" (p. 296). When a teacher explicitly explains her personal beliefs and experiences related to the literature, children are encouraged to connect the text to their lives, making reading useful and personal. A teacher may explicitly state processes that good readers naturally use. A teacher may find a word that she pretends not to know. In demonstrating a strategy to use visual cues, he or she may say out loud to the class, "Hmm, I am not sure what this says. I wonder if I can find clues on this page to help me know what this word might mean. Do you see anything that might help me?" This technique may be a useful strategy for struggling readers who may not know how to use visual cues to better understand a text.

Cambourne (2002) defines systematic instruction as "instruction that is based on proactive, rational planning. It is evidenced by formal planning documents that indicate the teacher has thought ahead and developed and documented a blueprint of future lessons, activities, resources needed, and assessment procedures that will be used" (p. 34). Cambourne argues that teachers must have a rationale for the methods that they use.

Being mindful is related to systematic instruction. Teachers must be mindful of their students' reading levels in order to provide a rationale for future instruction. Langer (1993)

suggests being mindful is having an "openness, not only to new information but to different points of view" (p. 68). Mindful teachers understand that the way their students learn a strategy will determine how they will use the strategy later. Therefore, teachers need to provide positive and motivating instruction. Constructivists are proponents of contextualized instruction that creates a need and purpose for reading. This view is in stark contrast to mindless, rote, strict instruction that relies heavily on repetition. Contextual instruction encourages the use of authentic texts, both fiction and non-fiction.

Instructional implications for teaching reading. During reading instruction, children are encouraged to increase their ability continually to reach higher cognitive levels of thinking. Adults are models of higher cognitive functioning and can be instrumental in providing scaffolding for children's growth. According to McGee and Richgels (2008), "conventional readers and writers have already mastered alphabet recognition and know most letter-sound associations. They now learn strategies for decoding words, meanings or new vocabulary words, and strategies for comprehending what they read" (p. 23). Emergent and conventional literacy are on a continuum of learning that varies for each individual.

Allington (2011) offers a few practical suggestions for educators. He recommends that large blocks of uninterrupted time be set aside for reading instruction. He recommends guided practice in small groups that is flexible. Intensive reading in small groups has proven to be beneficial to struggling readers (Phillips & Smith, 2010; Scanlon, Gelzheiser, Vellutino, Schatschneider, & Sweeny, 2010; Vellutino, Scanlon, Zhang, & Schatschneider, 2008). Groups of similar ability students may be better to tailor instruction for each group. However, Strickland, Ganske, and Monroe (2002) recommend also allowing opportunities for heterogeneous groups so that children avoid the stigma of being in a particular group. They also recommend that reading

instruction include multilevel activities to help ensure that struggling students will be able to engage with the task at their level. Strickland et al. (2002) encourage the teaching of self-monitoring. "They must be taught to question: Does this make sense? Does it sound right?" (p. 80).

Strickland et al. (2002) suggest that struggling readers treat reading like a problem-solving activity. In order to read and comprehend, a child must be taught different strategies to use. These strategies are taught explicitly and are scaffolded. First, a teacher models the use of the strategies. Next, a teacher makes explicit what he or she is thinking to solve a reading problem. Eventually, readers are expected to apply these strategies independently.

National Reading Panel Recommendations for Areas of Instructional Focus

In 2001, the National Reading Panel (NRP) issued a report with recommendations for effective reading instruction. The report reflected the results of an extensive literature review, of 300-400 research reports for each topic. In the report, the NRP concluded that effective reading programs include instruction in: (a) phonemic awareness, (b) phonics, (c) fluency, (d) vocabulary, and (e) comprehension. In order to become proficient readers, students need to develop skills in all five areas. Reading assessments for children often intend to assess each of these five areas.

Phonemic awareness. Phonemic awareness relates to the ability to hear and manipulate individual sounds in words. This includes the ability to segment, blend, delete, add, substitute individual sounds. For example, the word "cat" can be changed into another word (e.g. "hat") by substituting one phoneme for another. Phonemic awareness is a subcategory of phonological awareness. Both phonemic awareness and phonological awareness focus on the sounds of spoken words. Phonological awareness is broader in that it includes phonemic awareness and also the

ability of identify and manipulate larger parts of language (e.g. alliteration, rhyme, words, and syllables). Phonemic awareness is the ability to identify and manipulate individual sounds. Children demonstrate phonemic awareness of a spoken language, for example, when they can segment phonemes /d/, /aw/ and /g/ after hearing and saying the word dog" (McGee & Richgel, 2008, p. 399).

There is a strong correlation between reading success and phonemic awareness (Bradley & Bryant, 1983; Calfee, Lindamood, & Lindamood, 1973). Phonemic awareness helps children learn to read and spell words. The ability to aurally analyze the phonemic structure of words and the ability to retrieve essential information rapidly, such as letters and sounds, are strong predictors of reading success or failure (Torgesen, Wagner, & Roshotte., 1994). The NRP found that phonemic instruction significantly improved reading performance in three types of children: children progressing typically, younger children at risk for developing reading difficulties, and older children with a reading disability. The panel concluded that phonemic instruction is especially useful for improving reading for younger children who do not yet have strong phonemic awareness. Ehri and Nunes (2002) recommend that phonemic awareness be taught in small groups and that teachers model the use of phonemic awareness consistently in their instruction. These authors suggest that within existing activities, a teacher may be able to talk explicitly about the phonemic structure of a word.

Phonics. Phonics requires students not only to understand the different sounds in spoken language, but also to know and match letters or letter patterns with sounds, learn the rules of spelling, and use this information to decode (read) and encode (write) words (Chard & Osborn, 1999). Phonics skills are critical for reading success (Chard & Osborn, 1999; Duff, Hayiou-Thomas, & Hulme, 2012; NRP, 2001). Knowledge of phonics is important because it enables

students to develop a more or less reliable system for decoding new words (National Right to Read Foundation, 2000). Phonics is not always reliable because the English language has many words that are not regularly spelled. Although most researchers and teachers agree that learning phonics is effective, there is controversy in the most effective ways in teaching phonics. Cunningham and Cunningham (2002) report that explicit, motivating, and multileveled instruction is effective in teaching phonics. They have concluded that "any kind of well-organized and efficient phonics instruction is generally better than little or no phonics instruction" (p. 91).

Fluency. Fluency is the ability to read accurately and quickly and effortlessly with expression (Mather & Goldstein, 2001; NRP, 2001; Richards, 2000). Researchers agree that the ability to decode words accurately is essential for fluent reading (McGee & Richgel, 2008; Shinn, Good, Knutson, Tilly, & Collins, 1992). Children who have good word recognition skills are able to be effortless in their reading, thereby, focusing their cognitive energy and memory to comprehension and expression (Allington, 2011).

Learning to become fluent is difficult because many processes must take place at the same time. Samuels (2002) suggests that fluency is difficult for beginning readers because they have limited working memory space available for both decoding and expression. A learner goes through stages in word recognition skills. First, identifying a word is often not accurate. Next, a child may accurately decode a word, but the process takes time and is frequently not automatic. Finally, a student is able to effortlessly decode a word.

There are instructional strategies to help children become more fluent readers. Samuels (2002) recommends that teachers use motivating texts that are at an instructional level. Researchers strongly recommend repeated reading as a technique to improve reading fluency.

20

Repeated reading involves multiple readings of the same text. In a small group version of repeated reading children begin reading orally from multiple copies of the same text. After one to three minutes, the teacher calls "time" and children mark where they stopped with a pencil. After the reading, the teacher or the children discuss difficult words that they encountered. The children then reread the text and the process repeats two more times, with the children marking how far they progressed each time. Repeated reading of a text can improve word recognition, reading speed, and comprehension (Kuhn & Stahl, 2000, National Reading Panel, 2000).

Vocabulary. Vocabulary is knowledge of word meaning and usage. Vocabulary knowledge can be divided into two categories: (a) definitional knowledge and (b) contextual knowledge. Definitional knowledge is knowledge of a word in relation to other words. This category includes knowing synonyms and antonyms (Osborn & Armbruster, 2001). Contextual knowledge is knowledge of subtleties of meaning in different contexts. Researchers agree that knowing the meaning of words is strongly related to successful reading comprehension (Adams, 2001; Brabham & Villaume, 2002; Cunningham & Stanovich, 1997; Jenkins, Matlock & Slocum, 1989; Osborn & Armbruster, 2001; Penno, Wilkinson, & Moore, 2002).

Graves and Watts-Taffe (2002) summarize over 100 years of vocabulary research with six propositions:

1.   Vocabulary knowledge is one of the best indicators of verbal ability.

2.   Vocabulary difficulty strongly influences the readability of text.

3.   Teaching the vocabulary of a selection can improve students' comprehension of that selection.

4.   Growing up in poverty can seriously restrict the vocabulary children learn before beginning school and make attaining an adequate vocabulary a challenging task.

21

5.   Disadvantaged students are likely to have substantially smaller vocabularies than

their more advantaged classmates.

6.   Lack of vocabulary can be a crucial factor underlying the school failure of

disadvantaged students.

Vocabulary knowledge is important because it has been shown to be a major factor in

predicting reading comprehension difficulties (Beck, McKeown, & Worthy, 1995; Cunningham

and Stanovich; 1997). Researchers agree that "if the word-level processes are not mastered, it

will be impossible to carry out the higher order processes that are summarized as reading

comprehension strategies" (Pressley, Allington, Wharton-McDonald, Block, & Morrow, 2001, p.

551). Therefore, vocabulary knowledge is essential in order to comprehend a text.

Implications for the teaching of vocabulary are well documented. Kuhn and Stahl (2000)

found that vocabulary instruction was most effective when children are provided with contextual

and definitional information. He found that multiple encounters in a variety of contexts help

children learn new words. For teachers, Graves and Watts-Taffe (2002) recommend that teachers

encourage children to read intensively and expressively. Teacher selection of the books that

children read is important. If a book is too difficult, a child may become frustrated. A book too

easy may not provide opportunities to learn new words.

Comprehension. Reading comprehension is defined as the degree of understanding of a

text. Lyon (1998) reviewed 33 years of reading research, and after studying the reading

development, reported that text comprehension is commonly impeded by many factors:

(a) vocabulary deficits; (b) inadequate background knowledge relevant to the

information presented in text; (c) lack of familiarity with semantic and syntactic

structures that can be employed to predict and better understand word and

grammatical relationships; (d) lack of knowledge about different writing conventions that are employed by the author to achieve different purposes via text (humor, explanation, dialogue, etc); (e) lack of verbal reasoning ability which enables the reader to 'read between the lines'; and (f) the ability to remember and/or recall verbal information. (p. 4)

Historically, teachers believed that students must first become fluent readers before comprehending a text (Mancilla-Martinez, & Lesaux, 2010). Researchers have indicated that teachers can help students develop comprehension strategies that will increase their comprehension of a text (Block & Duffy, 2008; Duke & Pearson, 2002). Duke and Pearson (2002) recommend that teachers create supportive classrooms in order to teach students comprehension strategies. A supportive classroom would include a "great deal of time actually reading. As with decoding, all the explicit instruction in the world will not make students strong readers unless it is accompanied by lots of experience applying their knowledge, skills, and strategies during actual reading" (p. 207). They also recommend that teachers encourage students to read authentic texts, a range of text genres, texts rich in vocabulary, and high quality discussions about texts. Block and Duffy (2008) suggest that the following strategies be explicitly taught through teaching modeling and guided practice: predict, monitor, question, image, look-backs, rereads, fix it, infer, find main ideas, summarize, draw conclusions, evaluate, and synthesize (See Appendix A).

In summary, because reading includes a combination of these five skills, teachers rarely teach or assess these areas in isolation. Teaching or assessing a skill in isolation is not often appropriate for young readers because in primary grades instruction and assessment should include authentic activities, such as reading a book. Shavelson, Baxter, and Pine (1992) warn that

23

if we narrow the focus of our instruction and assessment in the early years and teach students only the skills necessary for them to achieve at high levels on standardized achievement tests, we are depriving them of the opportunity to lay the foundation for the higher order thinking skills necessary to become successful learners throughout their schooling.

Summer Reading Programs

Summer reading programs, as a way to support struggling readers, have become increasingly popular in the United States and are primarily used as a preventative measure to help children at risk for reading failure. Two types of summer school programs are the focus of the most recent summer school research: full-day programs and book access programs. Full-day programs are difficult to compare because they vary in many ways. Some programs employ certified teachers; others involve college interns or volunteers. Some use prescribed reading programs while others give teachers more freedom in curriculum decisions. The duration of the programs also varies. Researchers (Chaplin & Capizzano, 2006; Curry & Zyskowski, 2000; The Progress of Education Reform, 2009) determined the following research-based guidelines for full-day summer school programs:

1. Increase the duration and intensity.

2. Expand participation to all Title I students.

3. Use a balanced curriculum approach.

4. Provide incentives, including free lunch and breakfast.

5. Provide site-based programs.

6. Hire certified teachers.

7. Have small class sizes (no more than 14).

8. Provide professional development.

One example of an effective full-day program is the Student Opportunity to Accelerate Reading (SOAR) program created by Austin Independent School District in Austin, Texas (Curry & Zyskowski, 2000). This four-week program served 2,406 K-2 school students. Program teachers received two days of professional development and hands-on learning in balanced literacy. Students were assessed using the *Developmental Reading Assessment* (DRA). After the program, students had increased their reading ability on average 2.1 reading levels. This increase is equivalent to one-quarter to one-half of an academic year. Ninety-two percent of these students showed reading improvement. Thirty-six percent of students who began SOAR below grade level ended the program at or above grade level. Programs like SOAR provide evidence that summer reading programs may increase reading level.

Book-access programs are based on the theory that "the single summer activity that is most strongly and consistently related to summer learning is reading" (Hyens, 1978, p. 14). Research supports the hypothesis that summer reading loss can be attributed to the fact that low SES students do not have adequate access to books (Allington, et al., 2010). Therefore, one way to combat summer reading loss is to provide books for disadvantaged students. Book access is the common element in all the book-based programs. Hynes (1987) and Kim (2006) have made the following guidelines for book-access programs: (a) make self-selection of books a top priority (book fair model), (b) provide books for self-selection within the appropriate text levels, (c) purchase the books for the children, and (d) provide the program over multiple years to see results.

Although prior research suggests that summer reading programs had a significant positive effect on students' reading level in the short run, there is less evidence to suggest that this improvement can be sustained over a longer period. In a recent study of a book access program,

25

researchers randomly assigned 852 elementary students from high poverty schools into control and treatment groups (Allington et al., 2010). Students in the treatment group received 12 books at the end of the school year. These books were leveled at or just above a student's reading level and were selected by the student from a variety of genres and topics. These students received the books over a three-year period. The control group did not receive any books. By analyzing the scores on the reading portion of the state-mandated *Florida Comprehensive Achievement Test* the researchers found that there was a significant difference in scores between the two groups. The treatment group made reading gains during the summer, whereas the control group did not make gains. The Progress of Education Reform (2009) recommends that more research be done on long-term effects of summer reading programs.

Reading Assessment

Data from reading assessments are collected primarily for five types of decisions: referral, screening, classification and entitlement, instructional planning, and progress monitoring (Salvia & Ysseldyke, 1995). Reading assessments can be useful in determining if a child struggles with reading. Specifications in Guidance for the Reading First Program (U.S. Department of Education, 2005) required that educators in Reading First schools evaluate students in the five critical areas of reading instruction (i.e., phonemic awareness, phonics, vocabulary, fluency, and comprehension) as defined by the National Reading Panel.

Paris, Lawton, Turner, and Roth (1991) recommend that assessment be collaborative and authentic to promote learning and motivation. Teachers should be able to participate with students during assessments by questioning, hinting, prompting, and sharing. Teachers can observe when students self-correct, how they reason, when they are confused by misconceptions, and when problem-solving strategies. They also recommend that assessment be longitudinal.

Assessment that provides information about the child over time enables teachers to document each student's strengths and weaknesses and to focus instruction on student improvement. Finally, they recommend that assessment be multidimensional.

Genishi (1997) reports that young children are inconsistent in their day-to-day behavior and that, because of this inconsistency, paper and pencil tests made up of multiple choice items are confusing abstractions presented in an unfamiliar format. "The increased use of readiness and achievement tests…in the primary grades has presented children just in the process of becoming literate with testing material that is appropriate for conventional readers, those who are already literate" (Genishi, 1997, p. 62). Harlin and Lipa (1990) suggest that informal measures are better predictors of reading performance with young children than formal assessment measures. Unlike standardized readiness and achievement tests, which assess what children already know, informal assessments allow the teacher to determine student strengths and weaknesses in reading and to plan individualized instruction to meet the diverse needs found in a typical classroom.

Informal reading inventories. One type of reading assessment common for younger children is the information reading inventory (IRI). The purpose of an IRIs is to evaluate different aspects of students' reading performance. IRIs are based on notions implicit in developmental (Chall, 1983; Spear-Swerling & Sternberg, 1996) and interactive models of reading (Stanovich, 1980). IRIs typically leveled texts which are books with words and grammatical features with a variety of difficulty (Paris & Carpenter, 2003).

While a student reads a leveled text in an IRI a teacher takes notes of the miscues the child makes. After reading, teachers typically elicit a retelling and ask recall and comprehension questions. Each text the child reads is judged to be at his or her independent, instructional, or frustrational reading level based on two factors: the percentage of words the student read

correctly and the reader's level of comprehension. Other factors such as fluency, rate, or engagement may be taken into account. The National Association for the Education of Young Children defines developmentally appropriate assessment in primary grades as assessment that is appropriate for the age and experiences of young children, provides students opportunities to demonstrate their performance during authentic activities, and legitimately addresses not only what children can do independently, but what they can do with assistance from other children and adults (Bredekamp & Copple, 1997).

An IRI is an appropriate measure of reading ability because it estimates reading ability by having a student actually read instead of testing reading skills in isolation. IRIs are also developmentally appropriate for young readers. Although IRIs are typically used as an instructional diagnostic assessment, Nilsson (2008) suggests that IRIs can contribute valuable information to a school's instructional literacy program.

There are validity concerns with IRIs. Nilsson (2008) found the following concerns: (a) there is great variation in the way IRI text passages are structured (Quinn &Applegate, 2002), (b) while text passages generally become longer at the upper levels to align with the more demanding texts, across inventories passage lengths at the same levels vary, (c) graphics and pictures vary, and (d) comprehension questions vary in terms of which aspects of the text they centered on, as well as what dimensions, or levels, of comprehension they measured.

Reliability of scores obtained using IRIs is also a concern for both word identification and comprehension. Leslie and Caldwell (2006) provide data suggesting the forms for determining that reading comprehension levels may be used interchangeably. With respect to the alternate forms of the QRI text passages, Leslie and Caldwell found the reliabilities based on

comprehension scores were all above .80, and 75% of the reliability estimates were greater than or equal to .90. It is clear that more research must be done to determine the reliability of IRIs.

Standardized assessments. In contrast to IRIs, formal standardized tests provide information on students' reading ability by comparing individuals' performance to the mean scores of a norm group of students. Formal standardized tests are sometimes criticized because the questions are often in a multiple-choice format and do not include other assessment techniques. It may be difficult to measure a skill accurately, such as reading comprehension, using a multiple-choice test. For example, students may better show their reading comprehension skills by responding to questions about a reading passage with a written or oral task.

Standardized achievement tests are popular. The public, including legislators, teachers, and parents, have faith in quantitative comparisons. Standardized test scores give the perception that they are scientifically valid and rigorous while teachers' judgments are considered subjective and open to bias. Therefore, standardized tests appear to provide an unprejudiced picture of a child's ability. Standardized tests are not required in most states until students are in the third grade; however, some states are creating or using standardized tests in earlier grades in order to determine if children need early intervention before they take the mandated tests. Standardized tests have also received criticism in the ways they are used in education: (a) failure to consider measurement error, (b) the use of a single assessment as a true measure of a person's ability, (c) the use of a single score as the only criterion to make important decisions, (d) failure to provide specific information to be used to provide better instruction, and (e) the failure to recognize that a student's performance on an assessment is a complex system of many conditions (Nitko, 1996).

Currently, formal standardized assessments are criticized for their continued focus on narrow definitions of academic achievement of students. Standardized tests have been criticized

for young children because young children learn best and most by actively exploring their environment, using hands-on materials and building upon their natural curiosity and desire to make sense of the world around them (National Association of State Boards of Education, 1988,). Shepard (1994) reports that policy decisions driven by parental demands and accountability testing in higher grades produces a skills-driven curriculum in the primary grades. This developmentally inappropriate environment may consist of long periods of seatwork, high levels of stress, and an abundance of fill-in-the-blank worksheets.

Many of the tests are being used for purposes for which they were never designed or validated (Shepard, 1994). Tests that are valid for influencing classroom practice are not appropriate for making high-stakes decisions about individual children unless the curriculum, the teaching, and the tests are aligned (International Reading Association, 2002). High-stakes educational testing may be used to determine districts' funding, teachers' rewards and sanctions, and students' assignment to educational programs and is not compatible with the learning processes of young children and the instructional strategies used by early childhood teachers (Stiggins, 1995). With the push for standardized testing in the primary grades, teachers will feel pressured to teach the test or teach to the test. These tests "encourage teachers to focus on narrowly defined, isolated, surface 'skills,' and to spend disproportionate amounts of time in activities that promote the learning of these skills" (Chaillé, 2007, p. 74).

Traditional standardized assessments have also posed a problem for sociocultural theorists because they do not portray an accurate picture of a child's total literacy development because they do not provide information about a child's emergent literacy development at which instruction should be aimed (Dixon-Krauss, 1996). Vygotsky's sociocultural theory of education is an especially appropriate lens through which to look at the issue of assessment in primary

30

grades (Vygotsky, 1978). With the Vygotskian focus on learning through social interaction, it would appear that instruments that assess skills in isolation are giving a teacher information about what a child knows at a specific moment in time (i.e., the independent developmental level of the child) and fail to address the issue of what a child is able to do with teacher assistance (i.e., the potential developmental level of the child). Assessing a skill in isolation (e.g., selecting the correct phoneme or decoding a word) may not be appropriate in determining the reading ability of a child. Alderson (2000) believed that it is difficult testing skills in isolation because reading skills are interwoven. He wrote, "Isolating skills of readers also separates the readers from the nature of the text and the task associated with any reading activity does not follow the processes the test-taker engages in" (p. 304).

Designs Used to Evaluate the Effectiveness of Summer Reading Programs

Two designs useful in evaluating change in ability are the (a) pretest-posttest nonequivalent control group design and (b) multilevel growth modeling (MGM) design.

Pretest-posttest nonequivalent control group design. In the traditional pretest-posttest design, a group of persons are measured, a treatment is applied to them, and then they are measured again. An increase (or a decrease) of the group average from the first measurement observation to the second is ascribed to be the effect of the treatment. This design is represented as the following:

$$O_1 \quad\quad X \quad\quad O_2$$
$$\text{----------------------}$$
$$O_1 \quad\quad\quad\quad O_2$$

Whereas O_1 represents the pretest occasion, X represents the treatment, and O_2 represents the posttest occasion. In this design selection bias if presumed to be present, because treatment is not

randomly assigned. The rationale, advantages, and assumptions of the pretest-posttest nonequivalent control group design will be reviewed.

Rationale. The pretest measure is useful in describing score differences between the two groups. For example, a control group may outperform a treatment group. Differential selection may be observed from the differences in the two groups' pretest measure. In a pretest measure the control group may have had a significantly higher mean than the treatment group, confounding the results of the treatment.

Advantages. This quasi-experimental design useful when randomly assigning students to treatment or control groups is not possible or realistic. This is a particularly useful design in education, where random selection is not possible. For example, intensive intervention in reading would not be as necessary for children who are reading beyond their grade level. Instead, a school district may want to target children who are struggling readers and at risk for reading and academic failure. In this case the treatment group differs in reading ability from students that do not participate in the reading program.

Assumptions. Because groups are not randomly selected, there are many threats to validity. A strong assumption is that the two groups are not too different from one another. A strong assumption is that the smaller the difference on the pretest, the less likelihood of selection bias. In this sense a pretest can estimate group equivalency, unlike some randomly assigned group designs (Shadish, Cook, & Campbell, 2002).

Multilevel growth modeling. The pretest-posttest design has been widely criticized because so many influences other than the treatment may account for a change in scores from pre to post. In MGM the initial status, rate of increase, and the shape of the growth trajectory represent the primary parameters of interest. Glass, Willson, & Gottman (1997) describe possible

effects of an intervention on intercepts and slopes. It may abruptly change the level of the series, or change the level after a short delay; it may change the level of the series permanently, or only temporarily; the intervention may sharply, deflect a series formerly drifting downward, causing it to drift upward; it may make a highly variable series more stable, or vice versa. To complicate matters further, an intervention may work a combination of effects on a time-series, e.g. a downward drifting, highly variable series may show an abrupt change in level followed by a highly stable upward drift coincident with an intervention. (p.43) This section will review the rationale, advantages, assumptions, use of moderators, and use in educational research in MGM.

   Rationale. In traditional repeated measures analysis such as ANOVA, the effect of a treatment is determined by collecting data at fixed intervals. Repeated measures allow error variability to be examined within and between individuals. One disadvantage to this traditional design is that subjects with missing data points are deleted. Repeated measures ANOVA also assumes sphericity. This concept means that the variance of the difference in scores for each pair of time points is the same. However, longitudinal studies often violate this assumption because of the correlations among scores taken over longer periods of time (Arnold, 1992). MGM is useful in analyzing longitudinal data with repeated measures.

   Advantages. MGM has become increasingly popular because of its "apparent elegance in representing both collective and individual changes as a function of time" (Stoel, van Den Wittenboer, & Hox, 2003, p. 21). MGM is also known as hierarchical linear modeling (HLM) or random coefficient regression (RCR). One advantage of MGM is that it takes into account the multiple levels that exist in public schools (e.g., school, classrooms, and individual students). Examining the variation in outcomes that exist at various levels helps educators to examine different instructional variables (Heck & Thomas, 2000).

Another advantage to MGM is that observations across time are not required to be time-structured or to have balanced complete data. MGM is useful when longitudinal data are collected with repeated measures and the repeated measures are the first level of the multilevel model. The individual students are the grouping variable, and the repeated measurements are nested within the individual (Bickel, 2007). Cheung (2009) suggests that MGM is useful when you "want to test how intra-individual differences (within-subject variation over time or settings) and inter-individual differences (between-subject variation) are related" (p. 3). Another advantage of MGM is that interactions between variables can be modeled across different levels of analysis. Holt (2008) states,

> within a growth modeling framework, this allows for modeling the relationship between effects that are repeated measures (i.e., measured within-persons) and individual-level effects (i.e., measure at the person level). Covariates assessed at the person level are termed timer-invariant covariate, and analyst easily can incorporate them into the level-two equations of the multilevel growth model. (p. 118).

Recent studies (Armstrong, Dedrick & Greenbaum, 2003; Cusumano, Armstrong, Cohen, & Todd, 2006; Stipek & Miles, 2008) support the use of MLM as a method to demonstrate change in educational research. Specifically, Cusumano et al. (2006) explored the impact of early childhood educator training and coaching on literacy acquisition of preschool children. A three-level model was structured, examining within-child differences (reading scores), child characteristics (age, race, etc.) and classroom characteristics (treatment intensity, etc). McCoach, O'Connell, Reis, and Levitt (2006) utilized MLM to analyze the impact of school and classroom level characteristics on the reading growth of elementary school students.

34

Holt (2006) explains that multilevel growth modeling can separate growth trajectories for each individual. Hierarchical linear modeling (HLM) allows for intra-individual variability in growth, which in turn helps to explain inter-individual growth in achievement. Researchers (Arnold, 1992; Bryk & Raudenbush, 2001; Raudenbush, 2001) explain the advantages of using HLM:

1. It can explain student achievement and growth as a function of school-level or classroom-level characteristics, while taking into account the variance of student outcomes.

2. It can model the effects of student characteristics, such as gender, race-ethnicity, or socioeconomic status (SES) on achievement within schools or classrooms, and then explain the differences in these effects between schools or classrooms.

3. It can model the between and within school variances at the same time, and thus produce more accurate estimates of student outcomes.

4. It can produce better estimates of the predictors of student outcomes within schools and classrooms by borrowing information about these relationships from other schools and classrooms.

5. Growth curves may be different for each individual.

Assumptions. As with any statistical procedure, assumptions are required. Specific to MLM, Bryk and Raudenbush (1997) initially discussed the issue of normality, suggesting that both individual outcomes and growth parameters assume normal distributions. Whether this assumption holds in a specific situation can be assessed through examination of histograms (for outcomes) and outliers (for growth parameters). Covariance structure is the second assumption (Bryk and Raudenbush, 1997). MLM does not require identical data collection design for each

subject, rather, the flexibility of the model accepts varying numbers of data points and spacing between observations. Therefore, HLM uses a covariance structure that estimates model error variance at multiple levels. It considers random effects (Bryk & Raudenbush, 1997). Last, assumptions regarding the metric used to assess the outcome variable require that each observation be measured on a common metric to allow for change in growth across time as opposed to changes in the measurement scale

Arnold (1992) identified some concerns to take into account when applying growth modeling. One assumption with linear equations is that the errors are distributed normally and are independent of the variables in the equation. Normal distributions with equal variances are hard to guarantee across different levels. An assumption that the relationships are linear is often overlooked in MLM. Furthermore, the data must be hierarchical with enough cases within and between individuals to provide sufficient degrees of freedom for the linear equations. The data must be accurate and the measurement instruments must provide valid and reliable data because error at one level can lead to bias in the relationships at the next level.

Moderators. In addition to pre-post change, multilevel growth models can also be extended to include moderating factors that help explain individual variation in growth trajectory (Muthen, Khoo, Francis, & Boscardin, 2003). A major advantage of MGM is that it includes multiple pre-intervention and post-intervention observations. Multiple observations provide more stable, accurate, and reliable estimates of real intervention effects. MGM requires at least three testing occasions; however, more testing occasions will allow for more accurate estimation. MGM assumes that the residuals at different levels are independently and normally distributed. MGM allows for analysis of incomplete data if the data are missing at random.

36

Time-invariant covariates may include gender or race/ethnicity. In MGM the "data are a series of observations nested within the individual; therefore, the structure of the data can be person-specific and much more flexible" (Holt, 2008, p. 112). The basic multilevel linear growth model can assess both the initial status and linear change over time. These equations represent the unconditional linear growth model with random slopes and intercepts presented by Holt (2008):

Level 1 (Occasions):

$$Y_{tij} = \pi_{0ij} + \pi_{1ij} \; Time_{tij} + \varepsilon_{tij}$$

Level 2 (Students):

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + r_{1ij}$$

Level 3 (Schools):

$$\beta_{00j} = \gamma_{000} + u_{00j}$$

$$\beta_{10j} = \gamma_{100} + u_{10j}$$

These equations describe the model for $i = 1 \ldots n$ subjects in $j = 1 \ldots m$ schools across $t = 1 \ldots T$ testing occasions. At level 1, Y_{tij} is the value of the DRA2 score and *Time* is the centered and scaled value of time at occasion t for person i in school j. The growth parameters π_{0ij} and $\pi_{1ij,}$ respectively represent the intercept and linear rate of change for person i in school j, and ε_{tij} is the occasion-specific within-person residual not accounted for by the growth parameters. In this basic model, the first equation represents Level 1 and is the individual growth model and describes the outcome at time 1, the intercept and the rate of change of person i, and random fluctuations around the linear growth trajectory. The Level 2 equations describe the between-person variability in the growth parameters: the intercepts, π_{0ij} and the linear slopes, π_{1ij}. The

person-specific Level 2 residuals, r_{0ij} and r_{1ij} represent the random between-person differences in the Level 1 growth parameters π_{0ij} and π_{1ij}, respectively. The school-specific random effects in this model, β_{00j} and β_{10j}, represent the average intercept and average rate of growth, respectively, in school j. The Level 3 residuals, u_{00j} and u_{10j} represent the random between-school differences in the Level 2 school-level average growth parameter, β_{00j} and β_{10j}. Finally, γ_{000} and γ_{100} represent the estimated fixed effect estimates of the overall slope and intercept of *Time* across all students in all schools.

An example of a study that included moderating variables was done by McCoach, et al. (2006) who analyzed four waves of data from the Early Childhood Longitudinal Study— Kindergarten cohort (ECLS–K). The data had three levels: time, student, and school. By modeling the data they found that on average, students make much greater reading gains in first grade (2.65 points) than they do in kindergarten (1.67 points). They also found that student-level variables (including socioeconomic status, ethnicity, kindergarten entry age, and gender) were better able to explain between-schools variability in students' initial reading scores and students' reading growth than school-level variables (percentage of minority students, percentage of free-lunch students, and sector).

Chapter 3: Method

This chapter provides a description of the *Students and Teachers Achieving Reading Success* (STARS) program, setting, participants, and instrument. It will also discuss the research design and analyses performed that addressed the following questions:

1.   What effect does participation in the STARS program have on students' reading ability as measured by the Developmental Reading Assessment 2 between the first-grade spring administration and the second-grade fall administration?

2.   How does participation in the STARS program affect students' reading growth as measured by differences in their first grade reading growth trajectory and their second grade reading growth trajectory?

3.   To what extent is the relationship between first grade students' participation or nonparticipation in the STARS program and their post-treatment reading test score trajectories moderated by students' gender, socioeconomic status, or ethnicity?

Description of STARS

In 2000, literacy specialists at the Alpine School District (ASD) created STARS in order to mitigate the negative effects of summer reading loss of struggling readers exiting kindergarten, first, and second grade. STARS main focus is on exiting first-grade students. STARS administrators explain that "this quality time concentrates on basic skills, with some of Alpine District's finest teachers, changing the lives of students" (Alpine School District, 2011, p.1). Although Alpine School District teachers and administrators feel that STARS is an effective program, an evaluation of the effectiveness has yet to be performed.

Structure. STARS includes ten first grade STARS classrooms each year (with the exception of 2008 where there were eleven classrooms). Each year STARS includes only one

exiting kindergarten and one exiting second grade classroom. The program lasts for 20 school days approximately one week after the regular school session has ended. The program lasts three hours per day and focuses solely on reading and writing skills. There are no scheduled breaks for snacks or recess. STARS also includes a book access program. The book access program consists of a teacher sending home 3-5 books each STARS school day. The teacher provides a home reading worksheet where students write the title of each book they read at home. STARS teachers monitor home reading and contact a parent whose child is not reading at home. Although there are no descriptions of exactly how teachers teach in their classrooms, interviews with five STARS teachers who have taught STARS for more than four years suggest the following STARS schedule (Table 1).

STARS teachers. Teachers for the STARS program consist of regular ASD elementary teachers. There is one main expert teacher in each STARS classroom. This teacher is considered to be a literacy expert because they are a Reading Recovery© teacher and/or have a reading endorsement. Two additional certified elementary teachers assist in each classroom under the direction of the main expert teacher. Each main STARS teacher is given $150 to purchase leveled books for their students to keep after the STARS program. Each STARS teacher is provided with $250 for other classroom supplies.

Teacher training. STARS teacher training is four hours long and is generally held in April prior to the start of the program. All teachers receive credits towards mandatory professional development hours in order to keep their teacher's licenses current. Since 2007, training for STARS teachers has been conducted by an ASD literacy specialist who is an expert in Reading Recovery©. The content of teacher training focuses on the nature of explicit

instruction including training in specific instructional strategies to help students' improve their

reading ability.

Table 1

Typical STARS Schedule

Time	Activity
8:20-8:30	Singing in hall before class Students taking roll Write name to check in
8:30-9:00	Teacher reads a book Points out different sounds Relates book with students' lives Asks comprehension questions Rereads the book
9:00-10:00	Four 15-minute centers Matching letters to sounds Making rhymes Singing songs One on one reading Independent reading
10:00-10:30	Writer's workshop Write a thank you note to a volunteer Write a journal
10:30-11:20	Guided reading in small groups Rereading Shared reading
11:20-11:30	Take home five books to read

Explicit instruction. STARS teacher training includes information on providing

struggling readers with explicit reading instruction. STARS teachers are trained to include a

phonics-based method that includes explicit methods to connect letters with the sounds they

make. Torgesen (2005) states that explicit instruction "is instruction that does not leave anything

to chance and does not make assumptions about the skills and knowledge that children will

acquire on their own" (p. 5). Explicit instruction also requires that the meaning of words be directly taught and be explicitly practiced in order to make the words accessible to children as they read. Finally, explicit instruction also includes sequential instruction and practice in the use of comprehension strategies to help construct meaning. STARS teachers also take the role as a model, coach and, and scaffolder. For example, during training a teacher is taught to first model a reading strategy, such as looking through a book before reading it and guessing what is going to happen. Next, a teacher may guide a student or small group of students in practicing the strategy and eventually the student performs this strategy independently.

  Reading Recovery© instructional strategies. Specific explicit instructional strategies are also taught to the STARS teachers. The training is, in part, based on the 2005 book *Literacy Lessons Designed for Individuals* by Marie Clay. STARS teachers are taught the following explicit reading strategies: (a) self-monitoring, (b) cross-checking, (c) self-correcting, and (d) searching.

  Self-monitoring. Self-monitoring teaches students to stop, notice, and acknowledge moments of uncertainty when they are reading. Clay (2005) recommends that it is important at this stage that the child comes to check on his or her own understanding.

  Cross-checking. Students use cross-checking when they persevere in trying multiple strategies to decode a word. Clay (2005) states, "Cross-checking describes simple behaviors. The child learns that one kind of information can be compared with another kind of information." (p. 110)

  Self-correcting. Self-correcting happens when a child notices that they read something incorrectly and the child corrects themselves. Self-correcting may include rereading a passage for better understanding. Clay (2005) says,

The courage to make a mistake, the 'ear' to recognize that an error has occurred, the

patience to search for confirmation, these were the characteristics of children making

good progress in their first year of reading. . . . a child who was aware that 'something

was wrong' went back over alone or tried several responses until the error was corrected.

(p. 55)

Searching. Searching involves students searching for more information in order to

understand what they are reading. When students search they use both visual and invisible

information. "Ultimately, readers must build ways of searching for and using information in their

own heads, but you can teach for, prompt for, and reinforce behaviors in a way that supports the

process" (Clay, 2005, p. 78).

Setting

The ASD is located in the northern portion of Utah County, Utah. It includes 69 schools

and serves over 66,000 students, from preschool to twelfth grade. Students who attended a Title I

school in the first grade or have an Individual Education Plan were excluded from the study. For

this study, the central treatment under evaluation was participation (at least 75% attendance) in

the STARS reading program. The STARS program was available to ten schools in ASD every

year, with the exception of 2008, where 11 schools were selected. Table 2 indicates which

schools participated in the STARS program from 2007-2011.

Schools were selected if they were not a Title I school and if they were among the highest

schools in the district in terms of K-2 children who scored below grade level on the DRA2.

Consequently, the schools selected for the STARS program have differed every year. Each school

included one classroom of exiting first-graders. Six percent of the participants attend the program

Table 2

Participating Schools in the STARS Program by Year

School	Year				
	2007	2008	2009	2010	2011
A	A_1	A_2	A_3	A_4	A_5
B	B_1	B_2	B_3	B_4	B_5
C	C_1		C_3	C_4	C_5
D	D_1	D_2	D_3	D_4	
E	E_1	E_2	E_3		
F	F_1	F_2			
G	G_1	G_2			
H	H_1				
I	I_1				
J	J_1				
K		K_2	K_3	K_4	K_5
L		L_2	L_3		L_5
M		M_2		M_4	M_4
N		N_2			
O		O_3	O_3	O_4	
P			P_3	P_4	
Q			Q_3		
R				R_4	
S				S_4	
T					T_5
U					U_5
V					V_5
W					W_5

for more than one summer. Only three percent of students participated in STARS after both kindergarten and first grade. These three percent were not included in the analysis because of potential confounding effects.

Participants

This study relied on archival data of first and second-grade student DRA2 scores collected by ASD administrators. The ASD provided all of the data by querying their data base. (A detailed description of how the data were prepared is in Appendix B.)

ASD began tracking student DRA2 scores beginning in 2007. Data for this study begins with the school years 2006-2007 and continued through 2010-2011. DRA2 scores were collected from five cohorts of students (2006-2007 through 2010-2011) over their first and second grade. A *cohort* is a group of students who attended the first grade in the same year and within the same school.

Selection criteria for STARS participation. Since this study includes students who participated in STARS, ASD has given selection criteria for STARS participation. In order to participate in the STARS program, students must have demonstrated the following:

- score below 16 on the first-grade spring DRA2.

- attend a school selected to have the STARS program.

- be recommended by the STARS teacher at the child's school.

- have parent consent and commitment to have their child regularly attend summer school.

- not be on an Individualized Education Plan

- not have participated in STARS for after both Kindergarten and first grade.

Figure 1 describes the criteria used to decide if a student qualified to become a STARS participant.
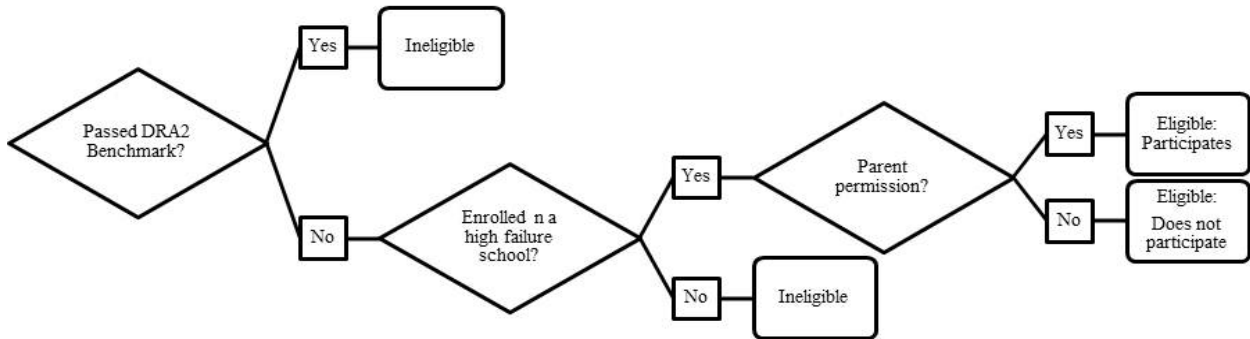
Figure 1. *STARS Participation Diagram*

ASD administers the DRA2 to all students three times a year in the first and second grade. Table 3 shows the benchmark and cap levels for the first-grade and second grade. For this study scores below 16 were considered not meeting the benchmark for the end of first grade. A *benchmark* is a standard against which individual student scores are compared. A benchmark is also known as a cut score. ASD has given teachers an artificial ceiling cap on DRA2 scores. This means that if an individual student reaches the cap score during testing, the DRA2 administrator stops the assessment even if the child's independent level may exceed the ceiling level.

Table 3

First and Second Grade Benchmarks and Caps for the DRA2

DRA2 Occasion	First Grade		Second Grade	
	Benchmark	Cap	Benchmark	Cap
Fall	3	16	18	28
Winter	10	20	24	28NF
Spring	16	24	28F	30F

Note: N = nonfiction, F = fiction.

Selection process for STARS participation. STARS teachers select STARS participants from those who have met ASD's selection criteria. STARS teachers met with their schools' regular classroom teachers in the school to discuss which students would benefit most from participating in STARS. Selection of students first focused on students with the lowest DRA2 scores. Although the selection procedure has not been well documented, data from informal interviews of 15 STARS teachers and the ASD STARS coordinator indicate that teachers may have also recommended children for the STARS program because they had high attendance throughout the school year and demonstrated good behavior that would encourage a cooperative learning environment for all students in the program.

After a student had been recommended for the program, their parents were contacted. Parents verbally committed to support their children in attending STARS by providing transportation and ensuring that their children would be able to attend STARS for the duration of the program. STARS participants who attended at least 75% of the STARS program were considered STARS participants for this study. Ninety-six percent of STARS participants met this requirement. The remaining four percent of students were not included in the analysis.

Not all students who were recommended to participate in STARS actually participated. Nonparticipation may have been because of the following: (a) inability to contact a parent, (b) language and communication barriers, (c) inability of a parent to provide transportation for the student, (d) family vacation plans, (e) unwillingness to commit to the STARS home reading program, (f) student resistance, or (g) illness of a student or parent.

Description of the nonequivalent groups. For this study, the researcher classified students into one of three nonequivalent groups. *STARS participant* refers to students who qualified for the STARS program by scoring less than 16 on the first-grade spring DRA2 and

participated in STARS the summer after first grade. *Eligible nonparticipants* refer to students who qualified for the STARS by scoring less than 16 on the DRA2, but did not participate in the program. *Ineligible nonparticipant* refers to students who did not qualify for the STARS by scoring 16 or higher on the first-grade spring DRA2, and did not participate. Table 4 indicates how many participants were included in the study for each of the three nonequivalent groups.

Table 4

Nonequivalent Group Participants

Nonequivalent Group	First Grade			Second Grade		
	Fall	Winter	Spring	Fall	Winter	Spring
STARS participant	789	815	820	815	814	805
Eligible nonparticipant	2150	2155	2326	2332	2178	2385
Ineligible nonparticipant	4780	4712	5079	5015	5048	5050

This method of classifying students into these three groups is consistent with the recommendations made by Battistin and Rettore (2002). They recommend that "every time an intervention is targeted to a population of eligible units but is actually administered to a sub-set of self-selected eligible units, it is worth collecting information separately on three groups of units: non-eligibles, eligible non-participants and eligible participants" (p. 13).

Table 5 describes the gender, ethnicity, and SES of the three nonequivalent groups. There was a slightly higher percentage of males in the STARS participants (51.4%) and Eligible nonparticipants (53.5%) compared to the Ineligible nonparticipant group (46.9%). The non-white ethnic group was (15.4%) comprised of all other ethnicities including Hispanic, Black, Asian, Polynesian, and Native American. SES was determined by students' participation in the school

free/reduced lunch program. ASD was unable to provide a complete record of participants' SES. The majority of the missing data for SES were in the school years for 2010 and 2011. These data were missing because SES data are most often collected by ASD when the state mandated *criterion-referenced test* (CRT) is administered in third grade. Students in first grade in the years 2010 and 2011 had not yet taken the CRT.

Table 5

Gender, Ethnicity, and SES of Nonequivalent Groups

	STARS participants		Eligible nonparticipants		Ineligible nonparticipants	
	n	Percent	*n*	Percent	*n*	Percent
Gender						
Female	429	48.5	1108	45.9	2107	50.5
Male	455	51.4	1292	53.5	1954	46.9
Missing	1	0.1	15	0.6	109	2.6
SES						
High	365	41.2	964	39.9	1980	47.5
Low	139	15.7	346	14.3	419	10.0
Missing	381	43.1	1105	45.8	1771	42.5
Ethnicity						
Non-white	136	15.4	305	12.6	269	6.5
White	748	84.5	2095	86.7	3792	90.9
Missing	1	0.1	15	0.6	109	2.6
Total	885		2415		4170	

Instrument

The DRA is an informal reading inventory that was developed in 1986 in the Upper Arlington City School District in Ohio by a committee of teachers and educators, headed by Joetta Beaver. Modeled after an informal reading inventory, the DRA is intended to be administered, scored, and interpreted by classroom teachers.

Purposes. The intended purpose of the DRA2 is to identify students' independent reading level, defined as a text on which students meet specific criteria in terms of engagement, oral fluency, and comprehension. The DRA2 attempts to provide an opportunity for reading to be assessed as a whole. However, specific measurements of isolated skills are required in the DRA2 scoring rubric (e.g. number of vocabulary words mentioned, length of phrasing, reading rate, etc.). Additional purposes include identifying students' reading strengths and weaknesses, planning instruction, and monitoring reading growth.

Other mentioned purposes of the DRA2 are to (a) diagnose students' instructional needs and plan for intervention as needed, (b) determine the level at which the student is able to read independently, (c) group students effectively to provide appropriate reading instruction and opportunities to practice reading skills and strategies, (d) document changes over time in reading performance by monitoring students'' ability to use a variety of skills and strategies, (e) identify students who may be working below proficiency and need further assessments, and (f) inform parents and other educators of students' current reading performance and achievement.

The five essential components of reading studied by the NRP are reported to be a focus of the DRA2 (Pearson Education, 2009). The DRA2 has been used in several studies to evaluate the effects of reading intervention programs in school districts (e.g., Donis-Keller, Saunders, Wang, & Weinstein, 2004; Curry & Zyskowski, 2000).

Advantages. The DRA2 is an informal reading inventory with authentic texts, instructionally relevant measures of fluency and comprehension, and results that are meaningful to classroom teachers, parents, and other stakeholders. One benefit of the DRA2 is that it documents reading growth over time. It is also intended to be diagnostic, so that teachers can determine what skills a child needs to focus on in order to become a proficient reader.

The DRA2 is a more natural and meaningful assessment compared to traditional tests because it is administered by students' teachers. This person is a familiar adult with whom children have become accustomed to interacting with making the DRA2 more developmentally appropriate for young students. The DRA2 can be considered an interaction between a teacher and students.

Materials. DRA2 includes a spiral-bound teacher guide, 23 leveled texts for kindergarten through first grade, a CD with copy masters of testing materials, a laminated assessment procedures overview card, a training DVD, a timing clipboard with calculator, an organizer with hanging folders, and 30 student assessment folders, packed in a storage box. The texts include full-color illustrations, which are meant to be motivating for students to read.

Scoring. Students' total score on the DRA2 determines if the text read is an intervention (frustrational), instructional, independent, or advanced level. The total score is a combination of three scores from each of the three main sections (engagement, oral fluency, and comprehension) that will be described. Each of these three sections has multiple subsections. A student is rated 1-4 in each subcategory. For example, 1 point is given for subsections reading performances that are considered to be at the intervention (frustrational) level, 2 = instructional, 3= independent, and 4 = advanced. In order to be considered at a particular DRA2 level a student must be considered independent in all three main sections. ASD has worked with Pearson Education to develop a Rubric Glossary of Terms to clarify how to score a DRA2 (See Appendix C). This glossary allows teachers to have more detailed information on how to score the DRA2.

Description of the constructs assessed. The DRA2 has three main sections: engagement, oral fluency, and comprehension. These three main sections along with their subcategories will be described. The criteria for DRA2 level 16 will be used in the following

examples. Level 16 is the level at which students are considered grade level in the spring of first grade.

Engagement. The engagement section requires the teacher to make a judgment on a student's past reading performance Engagement includes two subsections of *book section* and *sustained reading*. In order to be considered at an independent level in engagement for level 16 a total of 6 points must be obtained (See Table 6). For example, a student can score 3 points in both subcategories when a teacher notices that the student selects new texts with mostly no teacher support. A student must be able to identify their favorite book by the title and tell about specific event in that book. A student must also be also to sustain independent reading for 10-15 minutes at a time.

Table 6

DRA2 Engagement Criteria for Level 16

| Reading Level | Engagement | |
	Book Selection	Sustained Reading
Intervention (frustrational)	1 = Selects new texts with teacher support; uncertain about favorite book	1 = Sustains independent reading for a short period of time with much encouragement.
Instructional	2 = Selects new texts with moderate teacher support; tells about favorite book in general terms.	2 = Sustains independent reading with moderate encouragement.
Independent	3 = Selects new texts with mostly no teacher support; identifies favorite book title and tells about a specific event.	3 = Sustains independent reading for 10-15 minutes at a time.
Advanced	4 = Selects a variety of texts; identifies favorite book by title and gives and over-view of the book.	4 = Sustains independent reading for an extended period of time.

Oral fluency. Oral fluency score is determined by *expression, phrasing, rate,* and *accuracy*. The student is asked to read the book as the teacher follows along with a copy of the text. While the student reads aloud, the teacher uses a text-specific observation guide to record six types of errors: (a) substitutions, (b) omissions, (c) insertions, (d) reversals, (e) incorrectly sounded out words, and (f) words told by teacher. The teacher also writes notes about expression and phrasing, including long phrases and also pauses. The reading is also timed to determine rate of reading or the words per minutes (WPM) metric. In order to be considered independent at level 16 oral fluency a student must score 11-14 points (see Table 7). Phrasing is assessed only when a student knows the words. When a student sounds out a word, the phrasing assessment is not affected. In level 16 a child with 11 or fewer miscues has 95% or higher accuracy and is considered independent.

Table 7

DRA2 Oral Fluency Criteria for Level 16

Reading Level	Oral Fluency			
	Expression	Phrasing	Rate	Accuracy
Intervention (frustrational)	1 = No expression	1 = Mostly word-by-word	1 = 39 WPM or less	1 = 93%.
Instructional	2 = Little expression	2 = Short (1-3 words) phrases	2 = 40-59 WPM	2 = 94%
Independent	3 = Some expression	3 = Longer (4-7 words) phrases some of the time; heeds to most punctuation	3 = 60-70 WPM	3 = 95%-98%
Advanced	4 = Expression conveys meaning most of the time	4 = Longer (4-7 words) meaningful phrases most of the time; heeds to all punctuation	4 = 71 WPM or more	4 = 99%-100%

Repetitions, pauses, and self-corrections are not considered miscues when measuring *accuracy*. If the student's score for rate or accuracy falls in the emerging/intervention (i.e., lowest) range for that text, the assessment is stopped, and a lower level text is administered at another date. If a student has five or more miscues, the teacher completes an oral reading analysis, which involves copying each substitution and noting the number of miscues and teacher-supplied words and the types of decoding strategies and miscues. However, it appears that the DRA2 evaluates word identification using a total accuracy score in which all miscues are used to calculate word identification, not total acceptability in which selected miscues are used.

Comprehension. Table 8 describes that the comprehension score is based on ratings in these areas: (a) *previewing*, (b) *retelling: sequence of events,* (c) *retelling: characters and details,* (d) *retelling: vocabulary,* (e) *retelling: teacher support*, (f) *reflection*, and (g) *making connections*. The first subsection, *previewing* is assessed before the DRA2 leveled book is read. In *previewing,* a teacher a teacher introduces a book to the student and asks specific previewing questions. For example, a book may have a picture of two friends on the cover. A teacher may say, "In this book we will read about two girls. Look at the picture. What is going on in this picture?" The teacher will ask the student to look through the book and make predictions about what the book is about. A connection is described as a student connecting the book introduction with the pictures on the page as they preview the book. For example if a child sees a picture of a basketball on the book cover, they may connect that the story may be related to basketball or sports in general. The remaining subsections of comprehension are assessed after the text is read. At level 16 a student is considered independent in comprehension with a total score of 19-25. After students read the book and begin retelling, they cannot refer back to the book. ASD provided teachers with questions to ask their students in order to assess their comprehension

54

Table 8

DRA2 Comprehension Criteria for Level 16

	Comprehension						
	Previewing	Retelling: Sequence of Events	Retelling: Characters and Details	Retelling: Vocabulary	Retelling: Teacher Support	Reflection	Making Connections
Intervention	1 = Comments briefly about each even or action only when prompted	1 = Includes only 1-2 events or details (limited retelling)	1 = Refers to characters using general pronouns	1 = Uses general terms or labels, limited understanding of key words/concepts	1 = Retells with 5 or more prompts	1 = Gives unrelated responses	1 = Makes unrelated connections
Instructional	2 = Identifies and comments briefly about each even or action with some prompts	2 = Includes at least 3 events in random order	2 = Refers to some characters by name and includes some important details	2 = Uses some vocabulary from the text; some understanding of key words/concepts	2 = Retells with 3-4 prompts	2 = Gives limited response and/or general reason for opinion	2 = Makes a connection that reflects a limited understanding of the story
Independent	3 = Identifies and connects at least 3 key events without prompting.	3 = Includes most of the important events from the beginning, middle, and end in sequence	3 = Refers to most characters by name and includes some important detail	3 = Uses vocabulary from the text; basic understanding of most key words/concepts	3 = Retells with 1-2 prompts	3 = Gives specific story events/actions and a relevant reason for the response	3 = Makes a literal connection that reflects a basic understanding of the story
Advanced	4 = Identifies and connection with at least 4 key events without prompting.	4 = Includes all of the important events from the beginning, middle, and end in sequence	4 = Refers to all characters by name and includes most important detail	4 = Uses important vocabulary from the text; good understanding of key words/concepts	4 = Retells with no questions or prompts	4 = Gives a response and reason that reflects higher-level thinking	4 = Makes a thoughtful connection that reflects a deeper understanding of the story

55

ability. Important words are content vocabulary words that come from the text. To assess reflection a teacher may ask, "What part did you like the best?" To assess making connections a teaching may ask, "What did the story make you think of?"

Administration. The DRA2 has been administered by ASD since the 2005-2006 school year. The DRA2 is administered by a student's elementary school teacher in the fall (first 20 days of the school, winter (first 20 days after winter break), and spring (between April 15th and May 5th). In ASD the DRA2 is administered from kindergarten through sixth grade. In kindergarten, it is administered only in the winter and spring. There are no time limits in administering the DRA2. Teacher guides estimate 10 to 15 minutes to administer student reading survey for each student, 6 to 20 minutes for the one-to-one conference, and 30 to 45 minutes for the silent reading and written components, depending on level. Estimates are based on students who are reading on grade-level, however, and struggling readers are likely to require more time.

Reliability. The publisher of the DRA2, Pearson Education, claims that the DRA2 provides reliable and valid measures of a students' reading ability (2009). They claim that the DRA2 "determines each student's independent level with an evaluation of three components of reading: reading engagement, oral reading fluency, and comprehension" (Pearson Education, 2009, p. 1). Pearson Education reported the following reliability estimates for DRA2 oral reading fluency and comprehension scores: (a) internal consistency reliability ($\alpha = .542-.853$), (b) reading passage equivalency ($p < .05$), (c) test-retest reliability ($p = .93-.99$), and (d) inter-rater and expert rater reliabilities (.58-.89).

Williams (1999) provides evidence of inter-rater reliability for the DRA with a study involving 306 students in kindergarten through third grade reading on text levels from A to 44. Eighty-seven teachers in 10 states conducted and audio-taped DRA conferences with three or

56

more students, after which each tape was rated by a two other teachers. Inter-rater agreement based on Rasch analyses for five rating scale items (accuracy, comprehension, reading stage, phrasing, and reading rate) was .80 for the first two raters. When all three raters were considered, inter-rater agreement fell to .74. Internal consistency data collected during the Williams (1999) study cited above indicated high levels of consistency for the five items across all three raters (Cronbach's alpha = .98) and for DRA texts (.97).

Validity. The publisher also claims that the *DRA2* is a valid measurement of accuracy, fluency, and comprehension, as supported by the following validity evidence: (a) criterion-related validity, (b) construct validity, and (c) content validity. However, they caution that "like all assessments, it is a single source of evidence about a student's reading development. Instructional decisions are best made when using multiple sources of evidence about a reader" (Pearson Education, 2009, p. 10). The publisher reports that the DRA2 was designed to reflect the characteristics of good readers as reported in the research literature. The publishers of the DRA claim that it is based on Clay's *Observational Survey* (Clay, 1993), however, there is no evidence to support this claim. The technical manual also reports the results of teacher surveys (*n*s of 80 to 175) conducted after the field tests in which teachers responded to a variety of statements about the assessment materials, their utility, and other dimensions. Teachers agreed that the DRA was helpful in describing reading behavior and identifying instructional goals (Pearson Education, 2009).

In a study a study by Weber (2000), correlations between DRA K–3 independent reading level and grade equivalents for the comprehension subtest on the Iowa Tests of Basic Skills were generally in the moderate range (.54 to .84). For a sample of second grade students (*n* = 2470) from a large urban/suburban district in Fort Bend, Texas (Williams, 1999), DRA independent

level assessed at the end of the 1998-1999 school year was moderately correlated with fall of third grade normal curve equivalent (NCE) scores on the Iowa Test of Basic Skills Vocabulary and Reading Comprehension subtests and for Total Reading (rs = .68, .68, and .71, respectively).

Reliability and validity concerns. Many concerns have been raised regarding the reliability and validity of IRIs (e.g., Invernizzi, Justice, Landrum & Booker, 2005; Spector, 2005). Although efforts have been made to clarify administration and scoring procedures, the text selection process and many aspects of scoring on the DRA2 remain highly vulnerable to inconsistency. The DRA places high demands on teacher judgment in administration and scoring and may lack sufficient explanation to teachers on how to conduct these tasks accurately. To combat this concern, ASD has provided teachers with training on the administration and scoring of the DRA2, as well as provided literacy specialists in all schools who are DRA2 experts available to teachers.

Several aspects of the DRA2 are problematic and may compromise both reliability and validity. Instrumentation is a threat because a change in the administration in testing procedures can sometimes lead to inconsistent testing results. Shadish et al. (2002) are concerned that people "make changes in how records are kept or of how criteria of success and failure are defined" (pp. 179-180).

Rathvon (2004) reviewed the DRA and found five main concerns. First, text selection is based on teacher judgment rather than on an objective, standardized routing task. No theoretical rationale or empirical data are provided in support of this procedure. Teacher guides include tables that recommend that teachers use texts for assessment that are on, at, and above grade level. The procedure in deciding what is on, at, or above grade level remains highly subjective and vulnerable to the operation of confirmation bias (Nickerson, 1998).

Second, further inconsistency in the text selection process arises from a student choice component. After the teacher has selected three or four texts, students are then invited to choose a text that they prefer to read. Student selection of text is often not accounted for and may add variability.

Third, teacher guides indicate that the DRA can be administered over several days. These concerns are even more apparent when evaluating comprehension. Comprehension scores may differ for students who read the entire text and answer comprehension questions on the same day and those who complete the oral reading portion on one day and finish reading the selection and respond to questions on another day. A fourth concern regarding DRA administration procedures relates to the vague guidelines for word supply during the oral reading component. Although the record of oral reading guidelines included in the teacher guides indicates that a "word told by teacher" is an error, no information is provided as to when the teacher is to supply a word to a struggling student (e.g., after a 3-second pause, after a 5-second pause, after the student has made an attempt to decode the word, etc.). Differences in word supply procedures can have a significant impact on both reading rate and comprehension.

Finally, despite the publisher's assertion that the continua include "consistent, clear criteria" for scoring student responses, many aspects of scoring make the DRA highly vulnerable to inter-rater variance. Also, answers are not provided for comprehension items, all of which are open-ended. Therefore, teachers must be able to remember the content of the books well enough to score comprehension items on the 4-point rating scale, with small gradations among the four performance levels for many items. The ability to judge these differences may be especially difficult for new teachers with little experience administering the DRA2.

Using the DRA2 in program evaluation. Although the DRA2 is primarily used as an informal diagnostic measure, other researchers have used DRA2 scores in program evaluations. Buchanan (2002) used the DRA2 as the outcome variable of a summer school program in Louisiana. The DRA showed an increase in students' text levels across grades, as well as changes in DRA level for a matched sample of students (n = 32,739) over the three-year period. Curry and Zyskowski (2000) used the DRA as an outcome variable in evaluating the effectiveness of a summer school remedial reading program in six Austin, Texas school districts in 1999 and 2002 (n = 1,101 and 1,994, respectively). The study showed that the students in the program significantly increased their reading level compared to a nonequivalent comparison group**.**

Design and Analysis

This study consisted of two designs used to answer the research questions. The pretest-posttest nonequivalent control group design compared the differences between the groups (STARS participants and Eligible nonparticipants) at different times (pretest and posttest). Multilevel growth modeling was used to compare DRA2 score intercepts and slopes in reading growth trajectories and to estimate moderator impact on growth trajectories.

Pretest-posttest nonequivalent control design. A nonequivalent pretest-posttest design was used in order to answer the first research question dealing with the effect of participation in the STARS program on students' reading. The dependent variable for this study was the DRA2 text level. The independent variables were the groups (STARS participants or Eligible nonparticipants) as well as DRA2 testing occasion (pretest and posttest). This design compared DRA2 scores of STARS participants with DRA2 scores of Eligible nonparticipants before and after the STARS program. Ineligible nonparticipants were excluded from this design. This design

can be graphically below with STARS refers to STARS participation and ~STARS refers to nonparticipation:

$$O_{1S} \qquad STARS \qquad O_{2F}$$

$$O_{1S} \qquad \sim STARS \qquad O_{2F}$$

Data from DRA2 scores from the first grade spring testing occasion (O_{1S}) were used as the pretest; fall of second grade scores (O_{2F}) were used the a posttest

The two-way ANOVA design combined features of both between group and within-group designs. In this study students are nested within schools and within the treatment and control groups. The repeated measures aspect of the design was used to account for the correlation between students' scores on the pretest and posttest. The mixed ANOVA estimated (a) between group effects (STARS participants vs. Eligible non participants), (b) within group effects (pretest vs. posttest), and (c) interaction effects (impact of group assignment by Time). The equation for the mixed ANOVA is given below.

$$Y_{ij} = \beta_{0j} + \beta_1 T_{ij} + \beta_2 S_{ij} + \beta_3 T_{ij} S_{ij} + \varepsilon_{ij}$$

In this model, β_{0j} represents the intercept for school j, β_1 gives the effect for the posttest (T_{ij}), β_2 is the effect of participation in the STARS program (S_{ij}), and β_3 is the effect of the interaction between the posttest and program participation. The variance component for the school-level random intercepts (β_{0j}) was also estimated. In this study, it was of interest whether children who participated in the STARS program showed a more favorable pattern of change on the DRA2 between the pretest and the posttest than the nonequivalent control group. An unstructured error covariance matrix was used to account for the non-independence of the pretest and posttest measurements.

61

Multilevel growth modeling. MGM involves performing a hierarchically structured set of regressions. Although there are multiple steps involved in MGM, all steps occur simultaneously in the estimation procedure. MGM was used to estimate both the effectiveness of the STARS program on students' reading trajectories as well as the impact of moderators. The dependent variable for this design was the DRA2 score. The independent variable was the three nonequivalent groups (STARS participants, Eligible nonparticipants, and Ineligible nonparticipants) with three levels.

The second research question dealt with the effect of participation in STARS on students' reading growth trajectories. Level 1 in the MGM design refers to the repeated measures within students. Level 2 refers to the individual student level. Level was the school level. DRA2 scores from three different experimental (STARS participants, Eligible nonparticipants, and Ineligible nonparticipants) were analyzed.

In this design intercepts and slopes of DRA2 scores are compared among the three nonequivalent groups. MGM is represented in Table 9. Assessment occasion (or time) for DRA2 is represented as O; STARS refers to STARS participation; ~STARS refers to nonparticipants.

Table 9

Design Used in the Multilevel Growth Model Analysis

Nonequivalent Group	First Grade			Treatment Condition	Second Grade		
	Fall	Winter	Spring		Fall	Winter	Spring
STARS participant	O_{1F}	O_{1W}	O_{S1}	STARS	O_{2F}	O_{2W}	O_{2S}
Eligible nonparticipant	O_{1F}	O_{1W}	O_{S1}	~STARS	O_{2F}	O_{2W}	O_{2S}
Ineligible nonparticipant	O_{1F}	O_{1W}	O_{S1}	~STARS	O_{2F}	O_{2W}	O_{2S}

In order to answer the second research question regarding student reading trajectories, group differences in intercepts and slopes of DRA2 scores were compared. DRA2 testing occasions and scores were modeled at the Level 1. This model resulted in an equation for each individual that consists of regression estimating their DRA2 performance trajectory, in terms of an intercept and a rate of change. Reading trajectories for the three nonequivalent groups from the school years 2006-2007 through 2010-2011 were analyzed.

MGM was the design also used to answer the third research question dealing with the impact of moderating variables on the relationship between group and students' reading trajectories. Baron and Kenny (1986) describe a *moderator* as "a qualitative (e.g., sex, race, class) or quantitative (e.g., level of reward) variable that affects the direction and/or strength of the relation between an independent or predictor variable and a dependent or criterion variable.

In the more familiar ANOVA terms, a basic moderator effect can be represented as an interaction between a focal independent variable and a factor that specifies the appropriate conditions for its operation" (p. 1174). For this study, three time-invariant covariates were included in the model as potential moderators, namely: gender, socioeconomic status, and ethnicity. MGM allows the researcher to estimate the variation in growth patterns and the relationships with covariates both within and between individuals (Raudenbush & Bryk, 2002). Models were estimated treating each of the three designated demographic variables as moderators of program effects. Analysis to answer the third research question regarding moderation included time invariant covariates or predictors at Level 2 which were students' gender, socioeconomic status, and ethnicity. MGM determined if the trajectory (intercept and/or slope) differs between participant and nonparticipant groups. Level 1 is described as

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}(Time) + \pi_{2ij}(After) + \pi_{3ij}(Time*After) + \varepsilon_{tij}$$

The value Y_{tij} is the outcome variable which is the DRA2 score for student i at t (months since the DRA2 score in August preceding the second grade) within each school j. The parameter π_{0ij} represents the expected performance on the DRA2 midway between the spring of first grade and the fall of second grade. The parameter π_{1ij} represents the slope (growth rate) of the expected performance on the DRA2 for the student over the period of the study following treatment assignment. *Time* has been given a scale in one-year increments, such that the distance between fall of first grade and fall of second grade represents one year.

Time was then centered between the spring of first grade DRA2 testing occasion and the fall of second grade DRA2 testing occasion. Finally, ε_{tij} is the deviation of student i from his or her growth trajectory at time t in school j. *After* is a dichotomous variable indicating whether the assessment occasion was before (0) or after (1) the time when the STARS program took place in the summer. π_{3ij} (*Time*After*) is also included as the interaction between Time and After. A multilevel model estimating the effects of group membership (STARS participant, Eligible nonparticipant, Ineligible nonparticipant) and the interactions of group membership with the above variables were estimated. This model used an ante-dependence covariance structure for the student-level repeated measurements and school random effects for the appropriate intercepts.

Chapter 4: Results

This study examined the effectiveness of the STARS summer reading program. Effectiveness was estimated by between and within group by differences in reading ability as measured by the DRA2. This chapter includes information on descriptive statistics and results that answer the three research questions.

Descriptive Statistics

Descriptive statistics in the form of means and standard deviations were computed for each school. Means for each school are indicated in Table 10. The first-grade fall administration had a mean score of 5.12 and a standard deviation of .54. The first-grade spring occasion schools' mean DRA2 score varied from 16.13 to 19.33 with a standard deviation of 1.26. The standard deviations of the test occasions ranged from .54 to 1.52.

Differences in means can be seen from school to school. Some schools demonstrated increased DRA2 scores from the first-grade spring to the second-grade fall DRA2 administration. For example, Table 10 shows that students in school *C* had a mean DRA2 score of 14.44 in the first-grade spring assessment occasion. Students in school *C* had a mean increase score of 2.55 DRA2 text levels higher in the second-grade fall assessment occasion. Other schools, such as school *T* had a decrease in DRA2 scores after the summer break.

To simplify interpretation the means, schools A-L were graphed in Figure 2 and the means for schools M-X were graphed in Figure 3. These graphs illustrate that there was significant variability in DRA2 scores between schools.

Table 10

Mean DRA2 Scores by STARS School

School	First Grade			Second Grade		
	Fall	Winter	Spring	Fall	Winter	Spring
A	4.55	11.11	16.55	16.91	22.17	25.01
B	5.85	13.12	18.51	19.09	24.94	24.12
C	5.03	11.13	14.44	16.99	21.68	23.69
D	5.59	12.55	18.05	20.46	24.31	25.56
E	5.34	11.77	17.01	17.34	22.30	24.35
F	5.67	9.96	17.64	16.35	21.94	25.00
G	5.28	12.66	19.33	17.73	23.14	22.48
H	5.56	12.42	18.86	18.39	24.62	27.60
I	5.19	13.01	19.58	21.03	24.77	25.03
J	5.06	11.33	17.36	17.65	22.56	23.01
K	4.79	10.55	16.26	19.01	23.52	26.58
L	6.06	11.75	17.83	18.88	23.88	24.50
M	4.79	11.68	17.63	18.32	23.24	26.81
N	5.46	12.52	18.99	18.56	23.54	24.00
O	4.53	10.11	16.13	17.03	21.56	25.30
P	5.04	11.55	17.38	18.49	23.78	24.00
Q	4.70	11.78	18.45	18.84	23.88	26.99
R	6.12	12.17	19.23	18.83	23.80	26.42
S	4.65	10.95	16.77	16.13	22.36	25.65
T	5.07	11.79	17.22	16.90	22.26	26.80
U	4.62	11.90	17.68	17.69	22.89	23.56
V	5.38	12.76	19.04	20.26	24.13	21.78
W	3.89	9.42	16.13	16.85	23.09	26.53
X	4.62	10.49	16.65	16.95	22.83	25.00
Mean	5.12	11.60	17.61	18.11	23.22	24.99
SD	0.54	0.99	1.26	1.30	0.98	1.52

Note. SD = standard deviation.
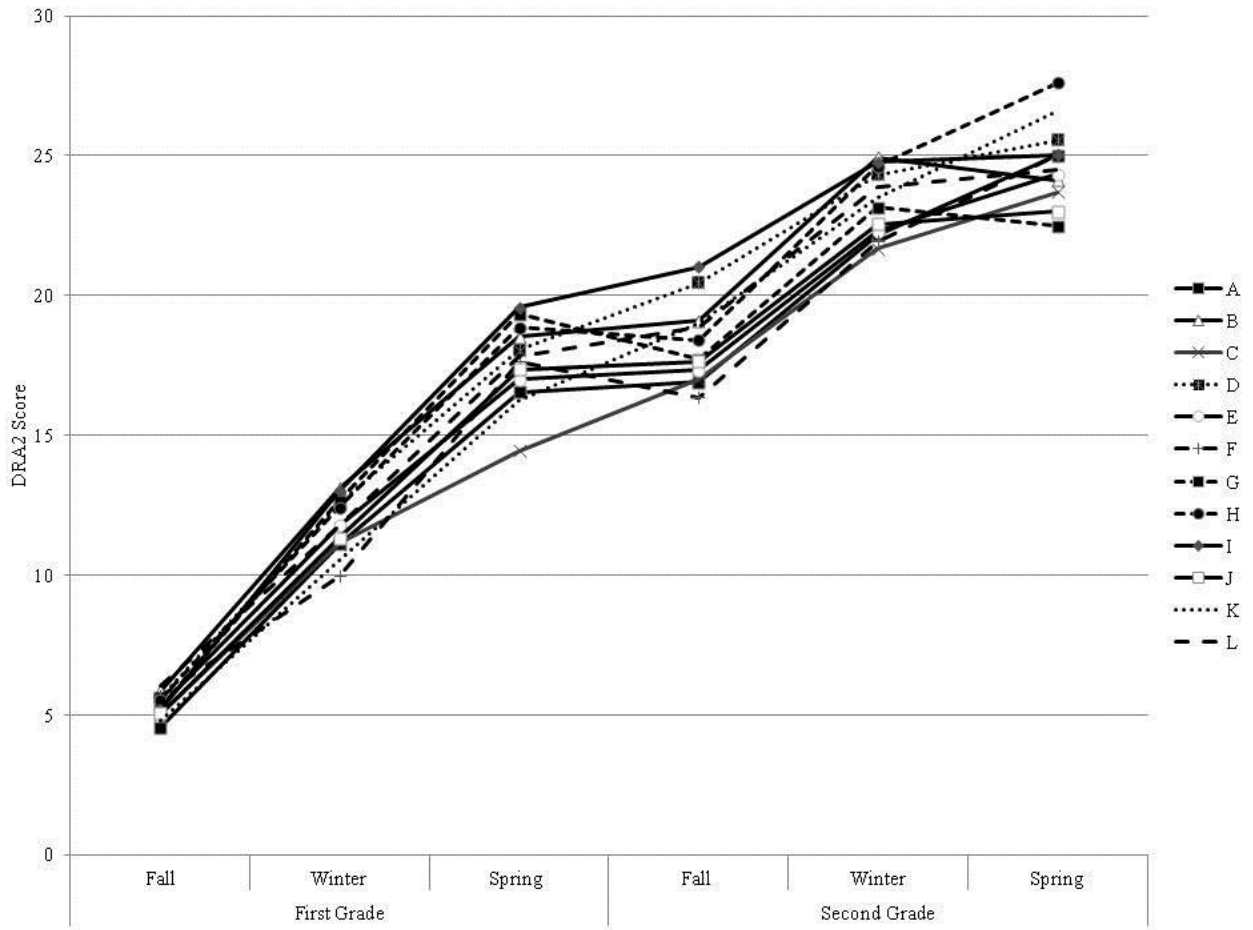
Figure 2. *Between-School Variation Schools A-L*
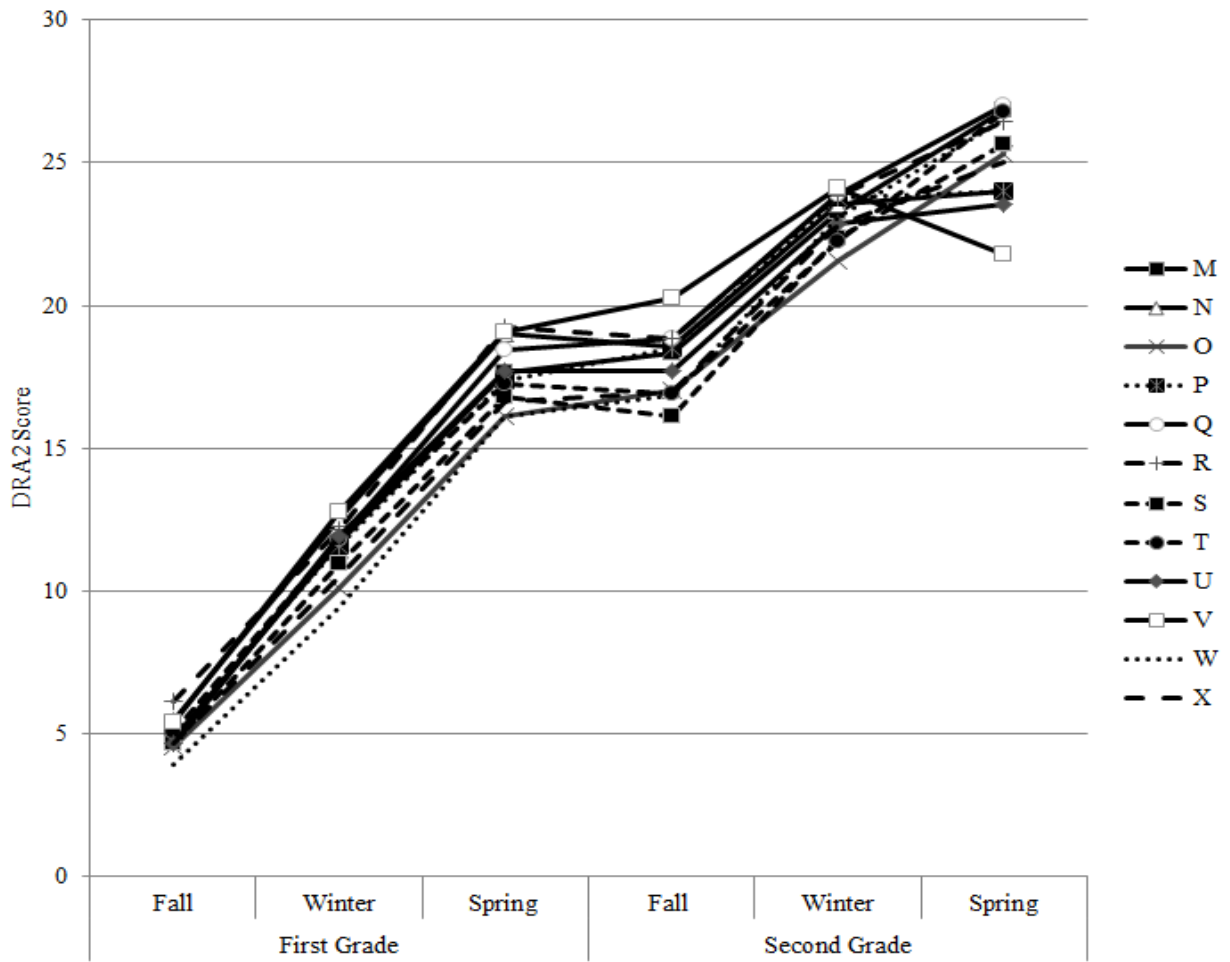
Figure 3. *Between-School Variation Schools M-X*

Table 11 described the mean DRA2 scores for first and second grade for all three nonequivalent groups. Ineligible nonparticipants had significantly higher initial DRA2 score mean (7.13) compared to the Eligible nonparticipants (2.68) and STARS participants (2.14). Mean scores for the STARS participants increased slightly between the first-grade spring occasion and the second-grade fall occasion.

STARS participants actually gained reading ability during the summer months as determined by the difference between the first-grade spring DRA2 mean score and the second-grade fall DRA2 mean score. The Eligible nonparticipants lost reading ability during the summer.

Table 11

Mean DRA2 Scores by Grade and Test Occasion

Nonequivalent Group	First Grade			Second Grade		
	Fall	Winter	Spring	Fall	Winter	Spring
STARS participants	2.14	5.59	10.78	11.21	17.31	22.13
Eligible nonparticipants	2.68	7.67	13.32	13.16	19.59	23.23
Ineligible nonparticipants	7.13	15.28	21.94	22.45	26.61	28.31
Grand Mean	3.98	9.52	15.35	15.61	21.17	24.57

Figure 4 graphically demonstrates that there was an overall positive growth trajectory in reading ability in the nonequivalent groups. The gap in reading ability is most pronounced at the first-grade spring test occasion as well as the second-grade fall test occasion. The gap decreases in the second-grade spring testing occasion.
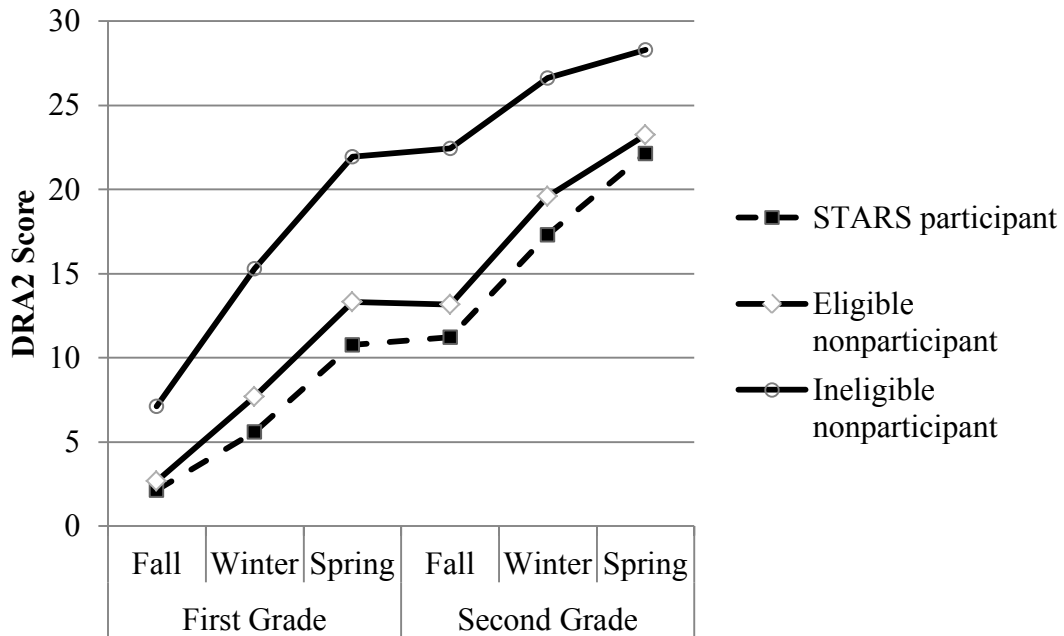
Figure 4. Mean DRA2 Score Trajectories for the Three Nonequivalent Groups

Mixed-Model ANOVA to Estimate Effect of STARS on Reading Ability

The first research question for this study addressed the effect STARS participation had on students' reading ability. The mixed-ANOVA focused on the difference between DRA2 scores on pretest and posttest within the two nonequivalent groups (STARS participants and Eligible nonparticipants). The spring of first grade was treated as a pretest and the fall of second grade was treated as the posttest measure. The Time variable is defined as the change in DRA2 scores between the pretest and posttest. The Group variable is defined as the difference between the two nonequivalent groups.

The means of the two nonequivalent groups are shown in Table 12. Figure 5 shows the means and confidence intervals graphically. STARS participants had an increase of DRA2 score between spring of first grade and fall of second grade. On average these students increase their

70

DRA2 score by .18 points. The Eligible nonparticipant group had an average decrease of DRA2

score by .44 points.

Table 12

Mean DRA2 Scores for STARS Participants and Eligible Nonparticipants

Group	Time	M	*SD*	df	95% CI LL	95% CI UL
STARS participants	First-grade spring	10.60	3.599	43.778	10.19	11.01
	Second-grade fall	10.78	4.542	78.935	10.31	11.56
Eligible nonparticipants	First-grade spring	13.31	3.829	26.785	12.94	13.68
	Second-grade fall	12.87	5.130	36.107	12.48	13.26

Note. CI = confidence interval; LL = lower limit; UL = upper limit.
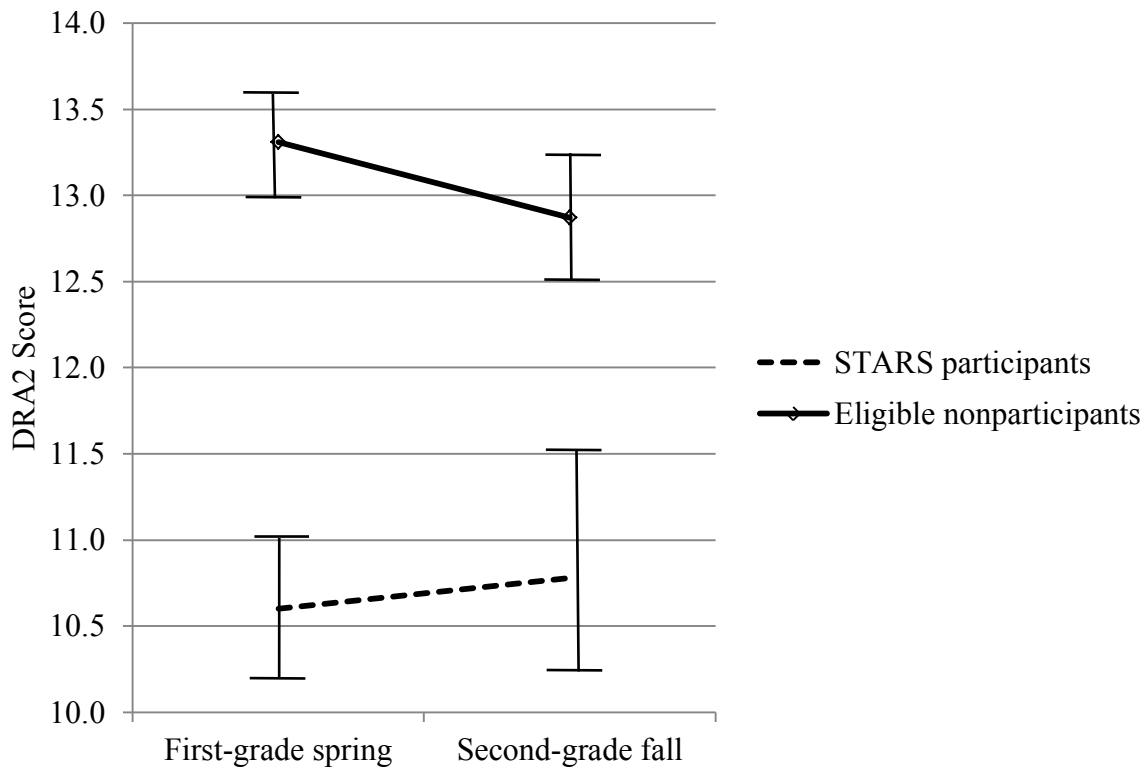
Figure 5. Mean DRA2 Scores for STARS Participants and Eligible Nonparticipants

Table 13 shows the main effects for Time and Group as well as the interaction of Group by Time. The *Group by Time* parameter estimate indicates that STARS participants had a statistically significantly higher increase in DRA2 scores compared to the Eligible nonparticipants from the spring of first grade to the fall of second grade, (*p* > .001). The groups functioned differently at the different time periods. While the STARS participants performed more poorly than the Eligible nonparticipants at the pretest, their performance increased at the posttest while the Eligible nonparticipants' performance decreased at the posttest.

Table 13

Mixed Model ANOVA Estimates for Two Nonequivalent Groups

Parameter	Estimate	SE	df	*t*	*p*	95% CI	
						LL	UL
Intercept	12.87	.192	27.255	66.885	.000	12.48	13.26
Group	-2.09	.200	3452.523	-10.405	.000	10.00	11.56
Time	.44	.081	2728.525	5.435	.000	12.76	13.86
Group by Time	-.62	.153	2731.630	-4.011	.000	11.56	12.94

Note. SE = standard error; df = degrees of freedom; CI = confidence interval; LL = lower limit; UL = upper limit.

The *intraclass correlation* (ICC) is defined as the degree to which individuals within a group or cluster are similar (O'Connell & McCoach, 2008, p. 4). Specifically, the ICC is the proportion of between group variance to the total variance. The higher the ICC, the more homogenous the groups are. The ICC estimates the degree of variance explained by the grouping structure in the population (Hox, Thomas, & Heck, 2010). The presence of a positive ICC estimate indicates dependency within the dataset. ICC is important to recognize because it

changes the error variance in single-level regression analyses. If there is little or no variation

estimated by the ICC, simple ordinary least squares regression analysis would be sufficient to

perform and a MGM would not be useful. Heck et al. (2010) present Level 2 and Level 3 ICCs

for a three level model as

$$\rho_2 = \sigma^2_{Level2} / (\sigma^2_{Level1} + \sigma^2_{Level2})$$

$$\rho_3 = \sigma^2_{Level3} / (\sigma^2_{Level1} + \sigma^2_{Level2} + \sigma^2_{Level3})$$

where ρ represents the ICC and σ^2 represents the between-group variance for each level of analy-

sis. Table 14 represents the between group variation necessary to calculate the ICC for level two

and three.

Table 14

Between Group Variance

					95% CI	
Between Group Variance	Estimate	*SE*	*WaldZ*	*p*	*LL*	*UL*
Level 1: Occasion	59.839	0.511	117.173	.000	58.847	60.849
Level 2: Student	15.162	0.486	31.190	.000	14.209	16.115
Level 3: School	4.377	1.620	2.702	.007	2.119	9.042

Note. SE = standard error; df = degrees of freedom; CI = confidence interval; LL = lower limit;
UL = upper limit.

The ICC for Level 2 (student) is 15.162 / (15.162+59.839) which is .202. The ICC for

Level 3 (school) is 4.377 / (4.377+ 15.162 + 59.829) which is .055. Both estimates are greater

than 0.05 which is a rough cut point set by researchers (Hox et al., 2010). Therefore, performing

multilevel analyses was appropriate for these data.

Multilevel Growth Modeling to Estimate Effect of STARS on Reading Trajectories

The second research question deals with how participation in STARS affects students' reading growth trajectories. Data from the six test administrations were included in this analysis. Differences in the intercept and slope of DRA2 scores between groups were determined. A dummy variable was created to indicate whether a testing occasion was before or after the STARS program. *Time* was centered between the spring of first grade and the fall of second grade testing administration. This analysis measures annual increase during both first and second grade.

Because repeated measurements are highly correlated, the ante-dependence covariance structure, introduced by Gabriel (1961, 1962), was used in analyzing the data. The ante-dependence covariance structure allows the variance between occasions to vary, allowing each observation to influence the subsequent observation. The ante-dependence covariance structure is in contrast to a standard first-order auto-regressive structure in which the error correlation is the same between all adjacent occasions and is therefore less flexible. The ante-dependence covariance structure is often a better fit to the data. Table 15 describes the twelve different intercepts and slopes for the three nonequivalent groups.

Reading loss was calculated by estimating the difference in reading ability of the following two predictive trajectories: (a) first-grade reading trajectory estimated by fall, winter, and spring DRA2 scores, and (b) second-grade reading trajectory estimated by fall, winter, and spring DRA2 scores. The expected mean DRA2 score for first graders in the Ineligible nonparticipant group in the spring of the first grade is 25.42. This group lost reading ability having a mean DRA2 score of 20.52 in the second grade. STARS participants also have an expected reading loss of 13.01 to 8.26. The Eligible nonparticipant group had similar DRA2

scores to the STARS participant group (16.09 to 11.13). These means indicate that STARS

participants lost slightly less reading ability on the average than the Eligible nonparticipants.

However, STARS participants lost slightly more reading ability than the Ineligible

nonparticipants. These means also indicate that STARS participants have a lower initial mean

than the Eligible nonparticipant and Ineligible nonparticipant groups.

Table 15

Separate Slopes and Intercepts of Three Nonequivalent Groups

						95% CI	
Parameter	Estimate	SE	df	t	p	LL	UL
Intercepts							
Ineligible nonparticipant before	25.42	.24	88.37	107.83	.000	24.95	25.88
Ineligible nonparticipant after	20.52	.23	139.63	89.96	.000	20.07	20.97
STARS participant before	13.01	.26	132.01	49.55	.000	12.49	13.53
STARS participant after	8.26	.30	242.67	27.99	.000	7.68	8.85
Eligible nonparticipant before	16.09	.24	98.16	66.33	.000	15.61	16.58
Eligible nonparticipant after	10.13	.24	168.15	41.59	.000	9.65	10.61
Slopes							
Ineligible nonparticipant before by time	21.90	.35	86.74	62.69	.000	21.20	22.59
Ineligible nonparticipant after by time	11.01	.35	144.23	31.67	.000	10.32	11.69
STARS participant before by time	13.15	.40	138.01	33.20	.000	12.37	13.93
STARS participant after by time	17.40	.46	247.08	38.23	.000	16.51	18.30
Eligible nonparticipant before by time	16.17	.36	99.25	44.61	.000	15.45	16.89
Eligible nonparticipant after by time	17.52	.37	170.19	47.10	.000	16.79	18.26

Note. SE = standard error; df = degrees of freedom; CI = confidence interval; LL = lower limit;
UL = upper limit.

Figure 6 depicts the intercepts and slopes of the three groups. The gray sections of the

lines in the graph represent predicted trajectory estimates of reading ability. These are the

expected DRA2 scores of the groups if they had continued in their first grade and second grade

trajectories. STARS participants on average lost less reading ability between the first and second grade compared to the other two groups. Figure 6 also shows that in the second grade, the gap in reading ability between STARS participants and the Ineligible nonparticipant group diminished substantially by the end of second grade.
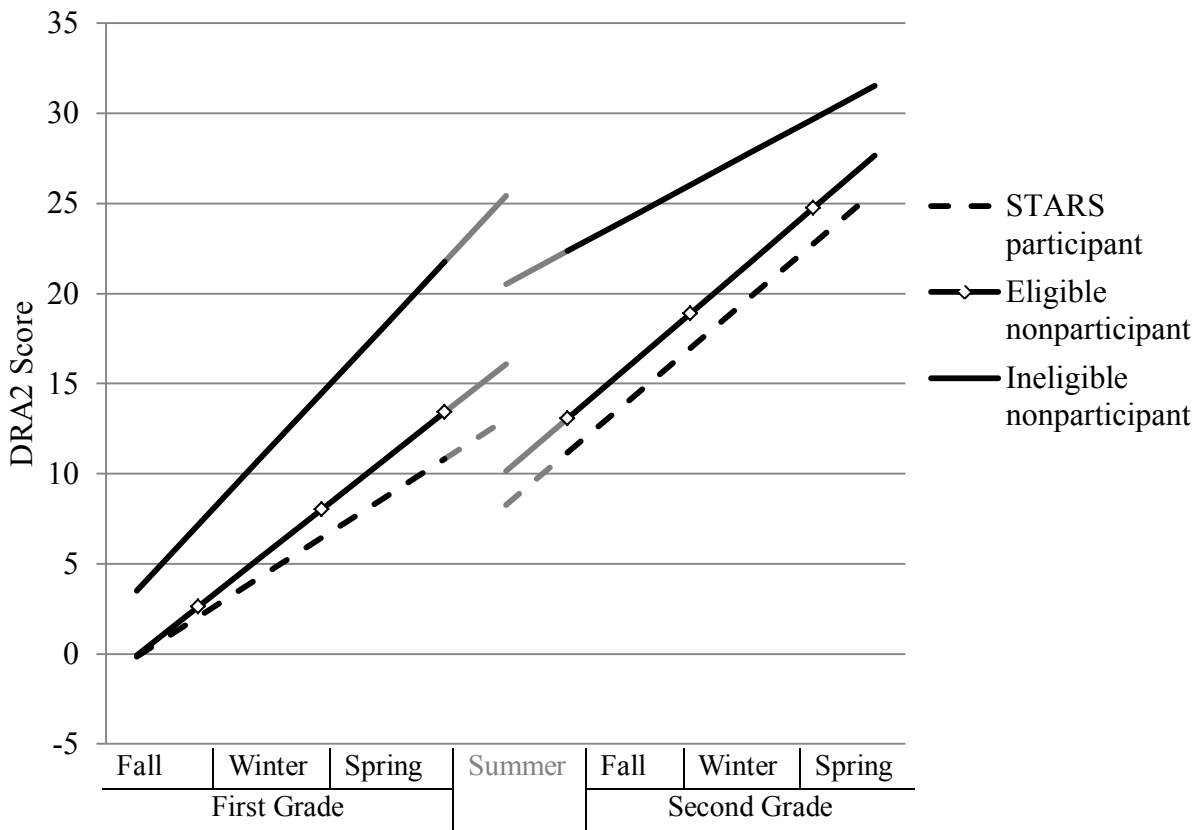

Figure 6. Results of the Multilevel Growth Modeling Analysis

Table 16 summarizes the differences in the intercepts and slope shifts between first and second grade for the three groups. Average summer reading loss was determined for the three groups. STARS participants lost less reading ability during the summer compared to the

Ineligible nonparticipant and Eligible nonparticipant groups. However, this figure indicates that STARS participants had similar slopes in reading ability in the second grade compared to Eligible nonparticipants. The average estimated summer reading loss for the Ineligible nonparticipant group was a loss of 4.90 points, 4.76 points for STARS participants, and 5.99 points for was 4.76 Eligible nonparticipants was 5.99. Group differences in reading loss were also analyzed. There is a nonsignificant difference of .15 points between the Ineligible nonparticipant and STARS participant groups ($p > .05$). The Eligible nonparticipants lost an average of 1.06 points more during the summer than the Ineligible nonparticipants ($p < .05$). The Eligible nonparticipants also lost more than the STARS participants (1.22 points, $p < .05$).

Table 16

Differences in the Intercepts and Slope Shifts of Nonequivalent Groups

Parameter	Estimate	SE	df	t	p	95% CI LL	95% CI UL
Summer Reading Loss							
Ineligible	-4.90	0.32	95.42	-15.42	.000	-5.53	-4.27
STARS	-4.75	0.36	124.37	-13.30	.000	-5.45	-4.04
Eligible	-5.96	0.33	103.95	-18.24	.000	-6.61	-5.32
STARS vs. Ineligible	0.15	0.48	110.14	0.32	.750	-0.79	-1.01
Eligible vs. Ineligible	-1.06	0.46	99.67	-2.33	.022	-1.97	-0.16
Eligible vs. STARS	-1.22	0.48	114.35	-2.51	.013	-2.18	-0.26
Before vs. After Slopes Shift Differences							
Ineligible	-10.90	0.50	113.09	-21.93	.000	-11.87	-9.91
STARS	4.26	0.61	201.17	6.93	.000	3.04	5.47
Eligible	1.36	0.53	133.1	2.58	.011	0.32	2.39
STARS vs. Ineligible	15.14	0.79	157.06	19.17	.000	13.58	16.71
Eligible vs. Ineligible	12.25	0.72	123.09	16.95	.000	10.82	13.68
STARS vs. Eligible	-2.90	0.81	167.26	-3.59	.000	-4.49	-1.30

Note. Ineligible = Ineligible nonparticipants; Eligible = Eligible nonparticipants; STARS = STARS participants.

Slope shifts were also analyzed in each of the three treatment groups. The initial slopes are calculated from the participants' reading trajectory in first grade compared to their reading trajectory in second grade. The Ineligible nonparticipant slope shifted downwards on average 10.89 points per year ($p < .001$) between the first and second grade. The slope of the STARS participant group shifted positively with an average of 4.26 points ($p < .001$). Finally, the Eligible nonparticipant group slope shifted the least at an average increase of 1.36 points ($p < .05$). Differences in the slope shifts between the three groups were determined. Slope differences between the STARS participant and Ineligible nonparticipant groups showed a significant difference of 15.14 points ($p < .001$). The Eligible nonparticipants also had a significant difference in slope shift compared to the Ineligible nonparticipant group (12.25 points, $p < .001$). The slope change for the STARS participants was 2.90 points more than the Eligible nonparticipants ($p < .001$).

Multilevel Growth Modeling of Moderator Effect on Reading Trajectories

The third research question addressed the impact moderators had on the relationship between STARS participation and students' reading trajectories. A MGM was performed with each of the moderators (gender, SES, ethnicity) treated as an additional fixed effect along with interactions between the moderator and the other model variables. This analysis included data from the same groups that were included to answer research question two: STARS participants, Eligible nonparticipations, and Ineligible nonparticipants. With respect to the moderation of program effects, the key tests in Table 17 are the interaction effects that incorporate the moderator together with the Group by After interaction term. *After* refers to DRA2 test occasion that occurred after the STARS program.

More specifically, the moderator by Group by After interaction tests whether moderation occurs for the group intercept changes, and the moderator by Group by After by Time interaction tests whether moderation is found for the group slope changes. None of the analyzed moderators had statistically significant effects on group differences in intervention-related reading test score slope differences. This is indicated by the absence of a moderator by Group by After by Time interaction for gender ($p = .339$), SES ($p = .284$), and ethnicity ($p = .287$).

Moderation of intervention-related intercept differences is indicated by a significant Moderator by Group by After interaction. These were non-significant for gender ($p = .110$) and ethnicity ($p = .379$), but the effect for SES ($p = .007$) was statistically significant. This latter effect indicates a difference between the students having reduced lunch status and the other students in terms of group differences in intervention-related intercept changes. With this possible exception, there is generally limited evidence of moderation of program effects on the reading trajectories by these selected demographic characteristics, with five of the six associated effects being statistically non-significant, even with a fairly large sample size.

As expected, gender, SES, and ethnicity each have significant independent effects ($p < .001$) on reading performance over the period of the study. As other studies usually find, students who are white, female, and have higher SES generally read better. There are also significant two-way interactions between group membership and each of these moderator variables ($p < .001$) across occasions. These between-subjects interaction effects, however, do not address program-related differences in reading trajectories.

Table 17

Moderating Variable Estimates

	Numerator df	Denominator df	F	p
Gender as Moderator				
Gender	1	7576.263	26.422	.000
Gender by Group	2	7596.289	13.765	.000
Gender by After	1	7532.619	.012	.912
Gender by Group by After	2	7565.810	2.207	.110
Gender by Time	1	12581.215	10.790	.001
Gender by Group by Time	2	12587.283	2.195	.111
Gender by After by Time	1	12420.388	.006	.941
Gender by Group by After by Time	2	12441.120	1.082	.339
SES as Moderator				
SES	1	4698.438	37.793	.000
SES by Group	2	4698.244	11.982	.000
SES by After	1	4801.872	.001	.979
SES by Group by After	2	4824.473	5.032	.007
SES by Time	1	7027.095	3.165	.075
SES by Group by Time	2	7033.255	14.077	.000
SES by After by Time	1	6649.274	.161	.688
SES by Group by After by Time	2	6726.578	1.259	.284
Ethnicity as Moderator				
Ethnicity	1	7908.832	52.282	.000
Ethnicity by Group	2	7911.350	10.836	.000
Ethnicity by After	1	7850.655	.701	.402
Ethnicity by Group by After	2	7863.656	.971	.379
Ethnicity by Time	1	12651.081	3.568	.059
Ethnicity by Group by Time	2	12704.941	10.777	.000
Ethnicity by After by Time	1	12858.180	.055	.815
Ethnicity by Group by After by Time	2	12882.618	1.249	.287

Other within-subjects interactions were included in order to retain the complete original model specification, but are also not relevant to hypotheses concerning program effects on reading trajectories. These include the non-significant moderator by After interactions for gender ($p = .912$), SES ($p = .979$), and ethnicity ($p = .402$) and also the nonsignificant Moderator-by-After-by-Time interactions for gender ($p = .941$), SES ($p = .688$), and ethnicity ($p = .815$). The moderator by Time interaction was statistically significant for gender ($p < .001$) but not for SES ($p = .075$) or ethnicity ($p = .059$), and the Moderator-by-Group-by-Time interaction was not significant for gender ($p = .111$), but was significant for SES and ethnicity ($p < .001$). Again, even the effects that are statistically significant have no bearing on the key study hypotheses about intervention-related reading trajectory changes. Given this, no attempt is made in this study to interpret these interactions, even though some of them may be statistically significant. In a study where no intervention was present, however, I recognize that studying these interactions and potential related hypotheses could prove useful.

Chapter 5: Discussion

When school is in session, most students progress in learning, although not all at the same rate. However, during the summer months, formal instruction usually stops for a few months and many of the gains that students had made during the previous academic year are lost over the summer break. The summer session between first and second grade is particularly problematic for students who do not have opportunities to read (Mraz & Rasinski, 2007). Unfortunately, summer reading loss significantly affects struggling readers who are already at risk for academic failure. Al Otaiba and Fuchs (2002) report that children experiencing reading difficulties in first grade remained poor readers in fourth grade with the gap between these children and their fluent peers widening over time. Studies have shown that reading gaps widen each year between first and sixth grade if students are not provided effective intervention (Helf, Konrad, & Algozzine, 2008).

The ASD developed the STARS program, an intensive summer reading program for struggling readers in order to mitigate summer reading loss. While STARS teachers and program administrators believe STARS has had significant positive effects on student reading levels there was a concern that the program may be vulnerable to budgetary cut. The purpose of this study was to evaluate the effectiveness of the STARS program. This chapter will compare findings in regard to the three research questions to existing literature. It will also discuss study insights gained, limitations, recommendation for practice, and suggestions for future research.

Summary of Results in the Context of Existing Literature

This section will review the results found in this study concerning the effectiveness of the STARS program. The results will then be compared to the findings of existing literature.

STARS effect on students' reading ability. This study found that STARS participants gained reading ability after the summer break, whereas Eligible nonparticipants lost reading ability. This difference was statistically significant ($p < .001$). This result is even more substantial given that STARS participants had significantly lower first-grade spring mean DRA2 score compared to Eligible nonparticipants. STARS participants gained reading ability through the summer months despite having lower initial mean score. Results of these analyses support previous studies that claim summer reading programs have statistically significant effects on the reading ability of early readers (Allington, 2003; Cooper et al., 2000; Chaplin & Capizzano, 2006; Curry & Zyskowski, 2000; The Progress of Education Reform, 2009).

STARS effect on students' reading trajectories. Educators and parents are often concerned about the long-term progress of individual children. The MGM results indicated that STARS participants lost less reading ability during the summer compared to the Ineligible nonparticipant and Eligible nonparticipant groups. STARS participants also had a higher slope shift in reading trajectories compared to the other two groups ($p < .001$). These findings support other claims made by previous research that indicate that summer reading programs can have a long-term effect on reading ability (Allington et. al., 2010; Curry & Zyskowski, 2000).

Moderator effect on STARS participation and reading trajectories. This study found that gender, SES, and ethnicity did not function as statistically significant moderators. These findings contradict what other researchers have found.

Gender. Previous research suggests that gender differences remain with gaps in reading favoring girls over boys (Dee, 2005) and that a larger number of males than females suffer from language disabilities such as dyslexia (Gleason & Ely, 2002). However, this study does not support the claim that girls perform better than boys in relation to STARS participation and

nonparticipation (Dee, 2005). However, the Gender-by-Group-by-Time interaction was statistically significant.

Ethnicity. Research indicates that although the achievement gap has narrowed somewhat, minority students are still at a disadvantage in U.S. schools (NCES, 2000). Differences in the experiences of minority children prior to entering kindergarten contribute to reading achievement differences in the primary grades. In this study ethnicity was not significant when included in the model.

Socio-economic status. According to Sirin (2005) family socioeconomic status is the strongest indicator of academic achievement. During the summer months, when school is not in session, children with low SES lose about two school months' worth of reading and math skills (Cooper, et al., 1996; Entwisle et al., 2005). There are a variety of reasons why summer reading loss is more prominent among poorer children. This study revealed that SES was not a moderating variable in reading ability in relation to participation or nonparticipation in STARS. One potential reason for the different findings regarding SES is that this study performed in suburban area, whereas, other studies showing SES differences are in urban settings (Sirin, 2005).

Insights Gained

Although the methods section of this dissertation includes the selection process of STARS participation, this information was initially not clearly understood by the researcher. Originally, a regression discontinuity design (RDD) was chosen to answer the first research question. When an RDD is used, students are assigned to a treatment or control group based solely on the location of their score relative to a predefined cut score (Lesik, 2006). Participants who have scores on one side of a cut score are given treatment, and scores on the other side of

the cut score are assigned to a control condition. The assignment variable is then subsequently used as a covariate in a regression analysis (Shadish, Cook, & Campbell, 2002).

Initially, interviews with STARS administrators indicated that STARS participation was based solely on the DRA2 cut score. Based on this information an RDD analysis was conducted producing a negative program effect. Further investigation of the data revealed that a large number of students met ASD criteria for STARS participation but were not selected to participate. Additional interviews with STARS administrators disclosed that STARS participants were not selected solely on a cut score, but that STARS teachers selected students based on a variety of other informal measures such as student behavior and attendance. STARS administrators were unaware of the substantial role STARS teachers had in the selection of STARS participants.

Because STARS participants were not selected based solely on a cut score RDD was determined to be an inappropriate design to answer the first research question. The insight gained in this experience was that it is essential that researchers, teachers, and administrators clearly communicate when collaborating. Another insight was that researchers must be vigilant in selecting the appropriate design for the given data and questions. In this case, selecting an inappropriate design produced inaccurate estimates potentially undermining the positive effect of STARS.

Limitations

As with other reading programs, it is difficult to ascertain which program component, collection of components, or other factors are responsible for positive outcomes. For example, the quantity of reading that children complete at home may prove to be the most influential piece

of the STARS program, regardless of the instructional method. A combination of quantity of reading and instructional method may also be influential in student reading improvement.

Limitations arise when the sole method of program evaluation is based on quantitative variables. Indeed, results from statistical analysis provide some evidence of program effectiveness; however, they often do not capture nuances and important information concerning program effectiveness. For example, individual interviews of parents and participants may have revealed a significant student increase in self-efficacy and identity as a reader. Smith (1988) argues that "ethnography rather than experimental psychology is the right horse for education to back" (p.123). She suggests "unlike experimental psychologists, cultural anthropologists have long recognized that it is impossible to study a situation objectively if investigators intrude their own rules, desires, or frames of reference" (p.120). Smith and other constructivists believe that statistical analyses are limited because they include data on a narrowly defined skill set. Assessments tend to show what a child knows, and does not offer information on how to better support a child's learning.

One other possible limitation to this study is that ASD has relatively low percentages of minorities (average 15% from 2006-2007 to 2010-2011). The finding in this study contradicts many previous studies (Allington, 2003; McNiece, Bidgood, & Sean, 2004) have found. Previous studies indicate that ethnic minority groups do not perform as well as whites in reading trends.

The degree of fidelity of implementation of the specified STARS instruction was not measured or included in the analysis in this study, hut it most likely varied from school to school and from year to year within a particular school. Teachers were provided with only four hours of STARS training, but no data were gathered to show how teachers' actually taught. However,

STARS coordinators have observed the teachers in their regular classrooms and have conferred with ASD principals about which teachers are held in high regard in teaching reading. Despite the training and assurances that STARS teachers are high quality, measurements on instructional approach were not collected.

The DRA2 has many potential limitations. Using the DRA2 as the outcome measurement to evaluate the effectiveness of STARS is problematic given that the DRA2 was administered 5-7 weeks months prior to STARS and 8-11 weeks after the STARS program. In these months prior to and after STARS students may have been involved in other reading programs, or have had other experiences influencing their reading ability. Records have not been kept on how DRA2 test administration has changed over the years and assessment fidelity was not measured in this study. Several aspects of the DRA2 are problematic and may compromise both reliability and validity including the following: (a) the DRA2 was administered by classroom teachers who had a wide range of training on how to administer the DRA2, (b) text selection was based on teacher judgment,(c) the DRA can be administered over several days, but no records were kept indicating to what extent the administration varied across days and from school to school, (d) vague guidelines for word supply during the oral reading component, and (d) many aspects of scoring make the DRA highly vulnerable to inter-rater variance.

Construct validity is of special concern with archival data. On the one hand, when using archival data the threats to reactivity are low because the participants do not know that they are in a study. Therefore, participants often do not alter their performance to manipulate a study's findings. On the other hand, construct validity is threatened when using archival data because the researcher is often forced to use outcome measures that are available, even if they are not fully

relevant to the treatment. This potential disconnect makes the available measures less sensitive for detecting a treatment effect.

Selection bias is a major concern with this study. Students were not randomly assigned. In order to participate in STARS, students not only had to fail to meet benchmark on the first-grade spring DRA2, but they also needed to be nominated by a teacher and have parental consent. This voluntary aspect of treatment creates selection bias since participation is, at least partly, voluntary. Data were not collected on the participant selection process; only anecdotal evidence suggests that many Eligible participants were not offered the opportunity to participate in STARS because of a variety of reasons such as behavior problems, low attendance, or higher DRA2 scores than other students that did not meet benchmark.

Future Research

This study focused exclusively on students moving from first grade into second grade, only two years of data. The results of this study could be extended if additional data were analyzed after subsequent summer recesses to see if the maintenance and gains found in this study persist over time with the cohort groups. Replicating the study to examine the impact of STARS on students moving from kindergarten to first grade and from second grade to third grade and beyond may offer additional evidence of the program's impact or lack of impact.

Different outcome measurements may be useful in determining the effectiveness of STARS. A future study could explore differences in scores on the state-mandated *criterion referenced test* for language arts between STARS participants and nonparticipants. STARS participation may also be correlated with later remedial or Special Education services. Graduation rates have been found to be correlated with early reading ability (Toppo, 2010). STARS participation may even have an effect on graduation rates. STARS may have been

effective in motivating students to love reading. Although students may not show immediate growth in reading ability determined by an IRI, some students may have acquired a love for reading that has later positive consequences. Student self-efficacy may also be affected by a positive experience in a summer reading program. Measures of student motivation and self-efficacy may be found to moderate the effectiveness of a summer reading program.

Propensity score matching may be useful in estimating program effectiveness (Rosenbaum & Rubin, 1981). In this design STARS participants would be matched to nonparticipants on a number of covariates (e.g. gender, SES, classroom teacher, DRA2 score). The use of propensity score matching should approximate random assignment of students to the treatment and control conditions thereby reducing selection bias and enabling the researcher may to make more valid causal inferences.

This study found significant differences among schools. A future study should examine the possible reasons for this school-to-school variability. Some schools may offer different types of summer reading programs that they have found to be successful. Also, there may be differences in how STARS is implemented in the various schools. Differential teacher effects may also help to account for school differences.

This study focused only on the use of gender, SES, and ethnicity as possible moderating variables. They were not found to be significantly related to the effectiveness of the STARS program. Previously mentioned statistically significant results were found in the moderators influence on group differences and time. An interesting study would be to find out more about how these moderators affect reading ability. Additional variables might also be considered in the future such as where students spend their summer (at home with a parent, in a formal daycare setting, with a babysitter, etc.) to better understand how summer instruction supplements

different summer childcare scenarios. Family structure might also be considered in future studies by examining the number of siblings in the home and whether the family has one or two parents.

Recommendations for Practice

Many scholars believe that educational research should emulate the scientific methods originally designed for the medical field. However, such an approach to finding the best way of teaching is unreasonable because each student, teacher, school, and context is unique. Experimental research often limits teachers to follow a strict method of instruction to ensure treatment fidelity. Random assignment into groups is often unfeasible in an educational setting. Insisting on strict instructional methodology in order to perform research minimizes the teachers' professional judgment and expertise, as well as their knowledge of their students. Research on how to best teach reading remains inconclusive. Any attempt to mandate a particular approach for all students should be seen as shortsighted, because one approach may not be effective for all students.

What is clear is that summer reading loss is a consistent finding in the U.S. Understanding the impact of the summer break on early readers is important because readers can experience significant reading loss due to the long summer break from instruction and lack of access to books. Reading loss is most pronounced among economically disadvantaged students and is an important, persistent issue that must be addressed (Bracey, 2002). This study addresses both the urgency of this need and the clear benefits that summer instruction can provide to students. Reading progress or loss over the summer has been linked to the number of books students read over a given period (Heyns, 1978, 1987), and the establishment of summer instructional programs with a reading emphasis provide a direct and fairly straightforward

opportunity to provide reading volume for students who participate. Providing students access to text can also be obtained by partnering with a local public library.

Where it is not possible to offer a comprehensive summer program, schools must find a way at minimum to provide all children access to books and, if possible, explicit instruction. Parents must be educated about the importance of reading over the summer months and encouraged to take an active role in reading with their children by providing instructional support with texts and offering incentives for participation.

School administrators should look to existing models for summer instructional programs and adapt them as appropriate to their own school setting. Though funding is a consistent challenge, the pursuit of funds for such initiatives is supported by strong statistical data that show the benefit of such programs. Summer reading loss accounts for at least 80 percent of the reading achievement gap by ninth grade, yet almost no federal or state programs or school district initiatives target summers as key to closing the achievement gap (Allington et al., 2010). School boards should use data from this study and those of other researchers as a basis for endorsing such efforts by their schools.

Summer reading programs should continue to be evaluated. If we are to achieve the NCLB goal of having all students reading on greade level by 2014, we must provide all students with access to learning opportunities. These opportunities should be provided during the summer, when children are at most risk for losing reading abilities. Effective summer reading programs, including access to books and explicit and motivating instruction, may mitigate summer reading loss for all students, regardless of their background. As a results, more children will graduate with reading skills necessary for success in life and in school.

References

Adams, M. J. (2001). *Beginning to read.* Cambridge, M.A.: The MIT Press.

Aikens, N. L., & Barbarin, O. (2008). Socioeconomic differences in reading trajectories. *Journal of Educational Psychology, 100*, 235-251.

Alderson, J. C. (2002). *Assessing reading.* Cambridge, UK: The University Press.

Allington R. L. (2011). What at-risk. *Educational Leadership, 3*, 40-45.

Allington, R. L., McGill-Franzen, A., Camilli, G., Williams, L., Graff, J., Zeig, J., Zmach, C., & Nowak, R. (2010). Addressing summer reading setback among economically disadvantaged elementary students. *Reading Psychology, 311*, 411-427.

Allington, R. L., & McGill-Franzen, A. (2003). The impact of summer setback on the reading achievement gap. *Phi Delta Kappan, 85*, 68-75.

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.

Alpine School District (2011). *STARS program*. Retrieved from http://www.alpine.k12.ut.us /asd/foundation/stars.html

Armstrong, K., Dedrick, R., & Greenbaum, P. (2003). Factors associated with community adjustment of young adults with serious emotional disturbance: A longitudinal analysis. *Journal of Emotional and Behavioral Disorders, 11*, 66-76.

Arnold, C. (1992). An introduction to hierarchical linear models. *Measurement and Evaluation in Counseling and Development, 25,* 58-90.

Bali, V. A., & Alvarez, R. M. (2004). The race gap in student achievement scores: longitudinal evidence from a racially diverse school district. *The Policy Studies Journal, 32,* 393-415.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social

psychological research: Conceptual, strategic, and statistical considerations. *Journal of*

Personality and Social Psychology, *51*, 1173-1182.

Bates, D. (2011). *White children will be in minority enrolling in U.S. schools by 2019, study*

finds. Retrieved from http://www.dailymail.co.uk/news/article-1355174/White-children

minority-enrolling-U-S-schools-2019-study-finds.html

Battistin, E., & Rettore, E. (2002). Another look at the regression discontinuity design. *Royal*

Economic Society Annual Conference. Cambridge, UK. Retrieved from

http://www.cepr.org/meets/wkcn/4/4528/papers/rettore.pdf

Beck, I., McKeown, M., & Worthy, J. (1995). Giving text a voice can improve students'

understanding. *Reading Research Quarterly, 30*, 220-238.

Bickel, R. (2007). Multilevel analysis for applied research: It's just regression! New York, NY:

 Guilford Press

Block, C. C., & Duffy, G. (2008). Research on teaching comprehension: Where we've been and

where we're going. In C. C. Block, & S. Parris (Eds.), *Comprehension instruction:*

Research based best practices (2nd ed., pp. 19-37). New York, NY: Guilford.

Bodrova, E., & Leong, D. J. (1996). *Tools of the mind: The Vygotskian approach to early*

childhood education. Columbus, OH: Prentice Hall.

Brabham, E. G., & Villaume, S. K. (2002). Vocabulary instruction: Concerns and visions. *The*

Reading Teacher, 56, 264-268.

Bracey, G. (2002). Summer loss: The phenomenon no one wants to deal with. *Phi Delta*

Kappan, 84, 12.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read: A causal connexion. *Nature, 301*, 419-421.

Bredekamp, S., & Copple, C. (Eds.) (1997). *Developmentally appropriate practice in early childhood* programs. Washington, DC: NAEYC.

Bryk, A. S., & Raudenbush, S. W. (1997). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*, 147-158.

Bryk, A. S., & Raudenbush, S. W. (2001). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Buchanan, T. K. (2002). *Developmental Reading Assessment: Student achievement.* New Orleans, LA: Louisiana Department of Education, Division of School Standards, Accountability, and Assistance.

Calfee, R. C., Lindamood, P., & Lindamood, C. (1973). Acoustic-phonetic skills and reading: Kindergarten through twelfth grade. *Journal of Educational Psychology, 64*, 293-298.

Cambourne, B. (2002). Holistic, integrated approaches to reading and language arts instruction: The constructivist framework of an instructional theory. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 25-47). Newark, DE: International Reading Association.

Catts, H. W., & Hogan, T. P. (2003). Language basis of reading disabilities and implications for early identification and remediation. *Reading Psychology, 24,* 223-246.

Chaillé, C. (2007) *Constructivism across the curriculum in early childhood classrooms: Big ideas as inspiration*. Boston, MA: Allyn & Bacon.

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

Chaplin, D., & Capizzano, J. (2006). *Impacts on a summer learning program: A random assignment study of building educated leaders for life (BELL)*. Mathematica Policy Research/The Urban Institute.

Chard, D. J., & Osborn, J. (1999). Phonics and word recognition instruction in early reading programs: Guidelines for accessibility. *Learning Disabilities Research and Practice, 14*, 107-117.

Clay, M. M. (1993). *An observation survey of early literacy achievement*. Portsmouth, NH: Heinemann.

Clay, M. M. (2005). *Literacy lessons: Designed for individuals, part two*. Portsmouth: NH: Heinemann.

Cooper, H., Chalton, K., Valentine, J. C., Muhlenbruck, L., & Borman, G. D. (2000). *Making the most of summer school: A meta-analytic and narrative review.* Boston, MA: Society for Research in Child Development.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greenhouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research, 66,* 227-268.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33,* 934-945.

Cunningham, P. M., & Cunningham, J. W. (2002). What we know about how to teach phonics. In A.E. Farstrup & S.J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 87-109). Newark, DE: International Reading Association.

Curry, J., & Zyskowski, G., (2000). *Summer opportunity to accelerate reading (S.O.A.R.)*
evaluation, 2000. Austin, TX: Austin Independent School District Office of Program
Evaluation.

Cusumano, D. L., Armstrong, K., Cohen, R., & Todd, M. (2006). Indirect impact: How early
childhood educator training and coaching impacted the acquisition of literacy skills in
preschool children. *Journal of Early Childhood Teacher Education, 27*, 363-377.

Dixon-Krauss, L. A. (Ed.). (1996). *Vygotsky in the classroom: Mediated literacy instruction and*
assessment. White Plains, NY: Longman.

Donis-Keller, C., Saunders, T., Wang, L., & Weinstein, M. (2004). *Second year evaluation*
report for the cornerstone literacy initiative. New York, NY: Institute for Education
and Social Policy, New York University.

Duff, F. J., Hayiou-Thomas, M. E., & Hulme, C. (2012). Evaluating the effectiveness of a
phonologically based reading intervention for struggling readers with varying language
profiles. *Reading and Writing: An Interdisciplinary Journal, 25*, 621-640.

Duke, N. K., & Pearson, P. D. (2002). Effective practices for developing reading comprehension.
In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading*
instruction (3rd ed., pp. 205-242). Newark, DE: International Reading Association.

Duncan, G. J., Dowsettb, C. J., Claessensa, A., Magnusonc, K., Hustonb, A. C., Klebanovd, P.,
Paganie, L., Feinsteinf, L., Engela, M., Brooks-Gunng, J., Sextonh, H., Duckworth, K., &
Japel, C. (2007). School readiness and later achievement. *Developmental Psychology,*
43(6). 1428-1446.

Durham Public School Office of Research and Accountability. (2001). *Results of the 2000-01 administration of the classroom observation survey/DRA as part of the DPS K-2* Student Assessment Portfolio. Durham, N.C.

Ehri, L. C., & Nunes, S., R. (2002). The role of phonemic awareness in learning to read. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (3rd ed., pp. 110-139). Newark, DE: International Reading Association.

Ely, R. (2005) Language and literacy in the school years. In J. B. Geason (ed.), *The development of language* (pp. 395-443) Boston, MA: Pearson.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2005). First grade and educational attainment by age 22: A new story. *American Journal of Sociology, 110,* 1458–1502.

Gabriel, K. R. (1961). The model of ante-dependence for data of biological growth. *Bulletin Institut International Statistique (Paris),* 39, 253-264 (33rd session).

Gabriel, K. R. (1962). Ante-dependence analysis of an ordered set of variables. *Annals of Mathematical Statistic, 33*, 201-212.

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.

Genishi, C. (1997, April). *Representing children in educational research: Locating Asian-American children in and out of classrooms*. Paper presented at the annual meeting of the American Educational Research Association. Chicago, IL.

Gersten, R. (1996). Literacy instruction for language-minority students: The transition years. *The Elementary School Journal,* 96, 228-244.

Glass, G. V. (1997). Interrupted time series quasi-experimental. In R. M. Jaeger (Ed.),

　　Complementary methods for research in education (pp. 589-608). Washington, DC:

　　American Educational Research Association.

Graves, M. F., & Watts-Taffe, S. M. (2002). The place of word consciousness in a research-based

　　vocabulary program. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say*

　　about reading instruction (3rd ed., pp. 140–165). Newark, DE: International Reading

　　Association.

Harlin, R., & Lipa, S. (1990). Emergent literacy: A comparison of formal and informal

　　assessment methods. *Reading Horizons, 30*, 45-55.

Heck, R. H., & Thomas, S. L. (2000). *An introduction to multilevel modeling techniques.*

　　Mahwah, NJ: Erlbaum.

Helf, S., Konrad, M., & Algozzine, B. (2008). Recouping and rethinking the effects of summer

　　vacation on reading achievement, *Journal of Research in Reading, 31*, 420-428.

Hoffman, K., & Liagas, C. (2003). *Status and trends in the education of blacks* (NCES

　　Publication No. 2003–034). Washington, DC: U.S. Department of Education.

Holt, J. (2006). How do we measure up? Capturing the complexities of educational growth. *Mid-*

　　Western Educational Researcher, 19(1), 2-7.

Heyns, B. (1978). *Summer learning and the effect of schooling.* New York, NY: Academic Press.

Heyns, B. (1987). Schooling and cognitive development: Is there a season for learning?

　　Child Development, 58, 1151-1160.

International Reading Association (2002). *Evidence-based reading instruction: Putting the*

　　National Reading Panel report into practice. Newark, DE: International Reading

　　Association.

Invernizzi, M., Justice, L., Landrum, T. J., & Booker, K. (2005). Early literacy screening in

    kindergarten: Widespread implementation in Virginia. *Journal of Literacy Research, 36,*

    479–500.

Jenkins, J., Matlock, B., & Slocum, T. A. (1989). Two approaches to vocabulary instruction: The

    teaching of individual word meanings and practice in deriving word meaning from

    context. *Reading Research Quarterly, 24,* 215-235.

Kagan, S. L., & Scott-Little, C. (2004). Early learning standards: Changing the parlance and

    practice of early childhood education? *Phi Delta Kappan, 85*, 388-396.

Kane, T. J., & Staiger, D. O. (2003). Unintended consequences of racial subgroup rules. In P. E.

    Peterson & M. R. West (Eds.), *No child left behind? The politics and practice of*

    accountability (pp. 152-189). Washington, DC: Brookings Institution Press.

Kim, J. S. (2006). *The effects of a voluntary summer reading intervention on reading*

    achievement: Results from a randomized field trial. Unpublished doctoral dissertation,

    University of California, Irvine.

Kuhn, M. R., & Stahl, S. A. (2000). Fluency*: A review of developmental and remedial practices*

    (CIERA Report No. 2-008). Ann Arbor, MI: Center for the Improvement of Early

    Reading Achievement. Retrieved from http//www.ciera.org/library/reports/inquiry-2/2-

    008/2- 008.pdf

Langer, E. J. (1993). A mindful education. *Educational Psychologist, 28,* 43-50.

Leslie, L., & Caldwell, J. (2006). *Qualitative reading inventory-4.* Upper Saddle River, NJ:

    Pearson Education.

Lesik, S. A. (2006). Applying the regression-discontinuity design to infer causality with non-

    random assignment. *The Review of Higher Education, 30(1),* 1-19.

Linnakylä, P., Malin, A., & Taube, K. (2004). Factors behind low reading literacy achievement. *Scandinavian Journal of Educational Research, 48,* 231–250.

Lyon, R. G. (1998). *The NICHD research program in reading development, reading disorders and reading instruction: A summary of research findings*. Special Education Programs (ED/OSERS). New York, NY: National Center for Learning Disabilities.

Mancilla-Martinez, J., Lesaux, N. K. (2010). Predictors of reading comprehension for struggling readers, *Journal of Educational Psychology, 102*, 701-711.

Mather, N., & Goldstein, S. (2001). *Learning disabilities and challenging behaviors: A guide to intervention and classroom management.* Baltimore, MD: Paul H. Brookes Publishing.

McAfee, O., & Leong, D. (2002). *Assessing and guiding young children's development and learning. (3rd ed.),* Boston, MA: Allyn & Bacon.

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

McGee, L. M., & Richgels, D. J. (2008). *Literacy's beginnings: Supporting young readers and writers.* Boston, MA: Pearson Education.

McNiece, R., Bidgood, P., & Sean, P. (2004). An investigation into using national longitudinal studies to examine trends in educational attainment and development. *Educational Research*, *46*, 119-136.

Mraz, M., & Rasinski, T. V. (2007). Summer reading loss. *The Reading Teacher, 60,* 784-789.

Muthen, B., Khoo, S., Francis, D. J., & Boscardin, C. K. (2003). Analysis of reading skills

    development from kindergarten through first grade: An application of growth mixture

    modeling to sequential processes. In N. Duan & S. Reise (Eds.), *Multilevel modeling:*

    Methodological issues and applications (pp. 71–89). Mahwah, NJ: Erlbaum.

National Association of State Boards of Education (1988). *Rethinking curriculum.*

    Alexandria, VA: Author.

National Center for Education Statistics. (2002). *The condition of education 2002.* Retrieved

    January 10, 2012, from http://nces.ed.gov/ pubsearch/pubsinfo.asp?pubid_2002025

National Reading Panel. (2001). *Teaching children to read: An evidence based assessment of the*

    scientific literature on reading and its implications for reading instruction. Bethesda,

    MD: National Reading Panel, National Institute of Child Health and Human

    Development. Retrieved from http://www.nichd.nih.gov/publications/nrp/smallbook. cfm

National Right to Read Foundation. (2000). *National reading panel reports combination of*

    teaching phonics, word sounds, giving feedback on oral reading most effective way to

    teach reading. Retrieved from http://nrrf.org/rdg_panel_rpt.htm

Neuman, S. B., & Celano, D. C., (2001). Access to print in low-income and middle-income

    communities. *Reading Research Quarterly, 36*, 8-26.

Neuman, S. B., Celano, D. C., Greco, A. N., & Shue, P. (2001). *Access for all: Closing the book*

    gap for children in early education. Newark, DE: International Reading Association.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review*

    of General Psychology, 2, 175-220.

Nilsson, N. L. (2008). A critical analysis of eight informal reading inventories. *The Reading*

    Teacher, 61, 526-536.

Nitko, A. J. (1996). *Educational assessment of students.* Englewood Cliffs, NJ: Prentice-Hall.

No Child Left Behind Act of 2001. (2002). Washington DC: U.S. Department of Education.

Retrieved from http://www2.ed.gov/policy/elsec/leg/esea02/index.html

Nyhan, R. C., & Alkadry, M. G. (1999). The impact of school resources on student achievement

test scores. *Journal of Education Finance, 25,* 211-227.

Oakes, J. (2005). *Keeping track: How schools structure inequality (2nd ed.).* New Haven, CT:

Yale University Press.

Osborn, J. H., & Armbruster, B. B. (2001). Vocabulary acquisition: Direct teaching and indirect

learning. *Basic Education, 46*(3), 7-11.

Paris, S. G., & Carpenter, R. D. (2003). FAQs about IRIs. *The Reading Teacher, 56(6),* 578-580.

Paris, S. G., Lawton, T. A., Turner, J. C., & Roth, J. L. (1991). A developmental perspective on

standardized testing. *Educational Researcher, 20(5),* 12-20.

The Partnership for Reading (2003). *A child becomes a reader: Birth through preschool.*

Retrieved from http://www.nichd.nih.gov/publications/pubs/birth_to_pre.cfm

Pearson Education. (2009). *DRA2: K-8 technical manual (2nd ed).* Upper Saddle River, NJ:

Pearson Education.

Penno, J. F., Moore, D. W., & Wilkinson, I. (2002). Vocabulary acquisition from teacher

explanation and repeated listening to stories: Do they overcome the Matthew effect?

Journal of Educational Psychology, 94, 23-34.

Phillips, G., & Smith, P. (2010). Closing the gaps: Literacy for the hardest to teach. In P.

Johnston (Ed.), *RTI in literacy: Responsive and comprehensive* (pp. 219–246). Newark,

DE: International Reading Association.

Pressley, M., Allington, R. L., Wharton-McDonald, R., Block, C. C., & Morrow, L. M. (2001). *Learning to read: Lessons from a first-grade classroom.* New York, NY: Guilford.

Progress of Education Reform (2009). *Summer Learning: Moving from the periphery to the core.* Denver, CO: Education Commission of the States.

Quinn, K.B., & Applegate, A.J. (2002). Levels of thinking required by comprehension questions in informal reading inventories. *The Reading Teacher, 56(2),* 174-180.

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

Raudenbush, S.W. and Bryk, A.S. (2002). *Hierarchical Linear Models* (2nd Ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology*, *52*, 501-525.

Ready, D., LoGerfo, L., Burkam, D., & Lee, V. (2005). Explaining girls' advantage in kindergarten literacy learning: Do classroom behaviors make a difference? *The Elementary School Journal, 106*, 21-38.

Richards, M. (2000). Be a good detective: Solve the case of oral reading fluency. *The Reading Teacher, 53,* 534-539.

Salvia, J., & Ysseldyke, J. E. (1995). *Assessment.* Boston, MA: Houghton Mifflin.

Samuels, S. J. (2002). Reading fluency: Its development and assessment. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction (3rd ed*, pp. 166-183). Newark, DE: International Reading Association.

Scanlon, D. M., Gelzheiser, L., Vellutino, F. R. Schatschneider, C., & Sweeny, J. M. (2010). Reducing the incidence of early reading difficulties: Professional development for classroom teachers versus direct interventions for children. In P. H. Johnston (Ed.), *RTI in literacy: Response and comprehensive* (pp. 257-291). Newark, DE: International Reading Association.

Shadish, W. R., Cook. T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin.

Shavelson, R. J., Baxter, G. P., & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. *Educational Researcher, 21*, 22-27.

Shepard, L. A. (1994). The challenges of assessing young children appropriately. *Phi Delta Kappan, 76(3),* 206- 212.

Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21*, 459-479.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75,* 417-453.

Smith, F. (1988). *Joining the literacy club: Further essays into education*. Portsmouth, NH: Heinemann.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Spear-Swerling, L., & Sternberg, R. J. (1996). *Off track: When poor readers become learning disabled.* Boulder, CO: Westview Press.

Spector, J. (2005). How reliable are informal reading inventories? *Psychology in the Schools, 42*, 593-603.

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16*, 32-37.

Stiggins, R. J. (1994). *Student-centered classroom assessment*. New York, NY: Macmillan.

Stoel, R. D., van Den Winttenboer, D., & Hox, J. (2003). Analyzing longitudinal data using multilevel regression and latent growth curve analysis. *Metodologia de las Ciencas del Comportamiento*, 5, 21-42.

Stipek, D., & Miles, S. (2008). Effects of aggression on achievement: Does conflict with the teacher make it worse? *Child Development, 79*, 1721-1735.

Strickland, D. S., Ganske, K., & Monroe, J .K. (2002). *Supporting struggling readers and writers*. Portland, ME: Stenhouse.

Toppo, G. (2010, May 18). Group links reading proficiency, national success. *USA Today*.

Torgesen, J. K., Wagner, R. K., & Rashotte, C.A. (1994) Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities, 27*, 276-286.

U.S. Department of Education. (2005). *The condition of education 2005 (No. NCES 2005-094)*. Washington, DC: National Center for Education Statistics.

Vellutino, F. R., Scanlon, D. M., Zhang, H., & Scatschneider, 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.

Vygotsky, L. S. (1978). *Problems of general psychology*. New York, NY: Plenum.

Weber, W. A. (2000). Developmental reading assessment and evaluation: A validation study. Retrieved from http://pearsonlearning.com/correlation/rsp/ResearchPaper_ DRA_Weber.pdf

White, S. B., Reynolds, P. D., Thomas, M. M., & Gitzlaff, N. J. (1993). Socioeconomic status and achievement revisited. *Urban Education, 28*, 328-343.

Williams, E. J. (1999). *Developmental Reading Assessment: Reliability study 1999.* Unpublished manuscript. Retrieved from http://www.pearsonlearning.com/correlation /rsp/ Research Paper_DRA.rtf

Appendix A: Effective Comprehension Strategies

Strategies that have been researched and validated to be highly successful since 2000

1. Predict—Size up a text in advance by looking at titles, text features, sections, pictures, and captions, continuously updating and repredicting what will occur next in a text.
2. Monitor—Activate many comprehension strategies to decode and derive meaning from words, phrases, sentences, and texts.
3. Question—Stop to reread and initiate comprehension processes when the meaning is unclear.
4. Image—Construct meanings expressed in text by wondering, noticing, and generating mental pictures.
5. Look-backs, rereads, and fix-it strategies—Continue to reflect on the text before, during, and after reading, continuously deciding how to shape the knowledge base for personal use.
6. Infer—Connect ideas in text based on personal experiences, knowledge of other texts, and general world knowledge, making certain that inferences are made quickly so as not to divert attention from the actual text but to help the reader better understand it.
7. Find main ideas, summarize, and draw conclusions—Make sure to include information gained from story grammar or textual features; if students can't make a valid summary of information read to date, this is the signal to go back to reread.
8. Evaluate—Approach a fictional text expecting to (and making certain that students do) note the setting, characters, and story grammar early on, with problems, solutions, and resolutions to occur thereafter.
9. Synthesize—Approach an informational text watching for textual features, accessing features, unique types of information, sequence of details and conclusions, and combining all of these to make meaning.

Appendix B: Data Preparation

To ensure anonymity of the students ASD de-identification the data. Initially, STARS participants were not included in the district's data base. Records of STARS participants were written on paper and placed in one binder for each summer from 2000-2011. These binders indicated which schools participated in the program each year, letters from parents indicating that they noticed reading improvement in their children, and student attendance records.

Unfortunately the binders including student participation and attendance records did not include student identification numbers, making it more difficult to locate the children in the data base. Therefore, the student names were manually typed into an Excel sheet and matched by the ASD data base. Eighty-five percent of the STARS participants' IDs were found in this way. Twelve percent of the STARS students were manually searched for matching their name, grade, and teacher. These students were not automatically found through the data base because of name misspellings and/or nicknames used instead of complete or correctly spelled names. For example, a child's name may have been written as Mike in the STARS binder; however his full first name was Michael. ASD data experts and I carefully matched students to their IDs by editing their names if the names were misspelled or nicknames had been used. If this was done, students were additionally matched on school and grade to ensure accuracy in assigning that student as a STARS participant. Three percent of STARS participants could not be located in the data base. The inability to locate all STARS participants may have been because of gross misspellings of student names or data entry problems. After all STARS participants were matched by ID, there STARS participation status was included in the ASD data base.

The list of schools participating in the STARS program changed somewhat from year to year. This information was also not included in the ASD data base. The STARS binders

Appendix B (Continued): Data Preparation

contained this information and the information was confirmed by the STARS director. A column was added to the Excel sheet provided by the school district indicating if a student attended a school that participated in the STARS program in the summer following that school year. This process was performed in Excel and checked for accuracy.

Data were placed into an SPSS data file and restructured by adding an "Occasion" column to represent each *DRA2* testing occasion per year. There were three testing occasions for Grades 1 and 2. This occasion variable was used to create the appropriately scaled and centered value of time as described in the Method section. After reviewing score frequencies by occasion, the counts indicated that some scores were above the maximum score allowed by the *DRA2*. The maximum score rule was enforced by explicitly designating caps in *DRA2* text level scores.

Students were included in the analysis if they attended a school that participated in the STARS program in the summer after their first grade. Inclusion of only students within a school that had the STARS program for that year was performed because students are assumed to be more alike to students within the same school. Because STARS schools changed from year to year, students who qualified for the analysis also changed each year.

ASD had data on gender, SES, and ethnicity. These data were converted to numeric values. Gender was coded female = 0 and male = 1. Socioeconomic status, as determined by reduced or free lunch status, was coded as non-low socioeconomic status = 0 and low-socioeconomic status = 1. Ethnicity was coded as White = 0; non-White = 1.

DRA Rubric Glossary of Terms

Right to Left Use of DRA Rubrics: When assigning a rating from the rubric ALWAYS begin with column #4. Say yes or no. If you say no, look at the third column. Say yes or no.
Never ask: Is it more like a 2 or a 3?

Most: "most" means half or more for any trait in the rubrics

ORAL READING/FLUENCY

Phrasing: This can only be determined on the portions of the text that are correct reading. Any time a student is working on a word to solve it should not be used as part of phrasing

•**Longer Meaningful Phrases:** 5-7 words at a time or the reading follows natural punctuation

•**Short Phrases:** 2-4 word phrases

Use of Cues (Levels A-3): While reading text the student uses meaning, structure, visual, and/or repeated patterns

Monitoring/Self-Correction: Accurate reading doesn't mean 100% reading in the context of self-corrections. The expectation of self-correction is to correct once for every three mistakes made, so 2 errors are still considered accurate for self-correction standards

Problem Solving Unknown Words: "At difficulty" means at the point of an error, self-correction or re-reading, are they using more than one type of cue

CONCEPTS OF PRINT (levels A-3 only)

Directionality: "Controls" means no lapses

One-to-One Correspondence: One thing they say is one thing they touch. Not a sliding motion.

Words/Letters: The teacher uses masking cards to isolate the requested letter. Not just pointing.

COMPREHENSION

Previewing (level 3-16): "Connects" means connects the teacher's book introduction with the pictures on the page as they preview the book

Questioning and Prediction (level 40-80):
        •"Reasonable" means connected to the text. "Thoughtful" means additional inferential explanation is given beyond connecting to something specific in the text.
        •"Beyond the text read aloud" means after previewing and reading for the running record that question/prediction has not already been answered. It may be based on something they saw in the book that will come after the oral reading part and that still counts.
        •"Three ?s and predictions" means 3 total, not 3 of each. A question and a prediction can be about the same thing and still count as separate answers.

Fiction Use of Text Features: Important characters are directly involved in the story structure elements, Main character and recurring support character(s). 3 details total, not 3 for each character.

Non-Fiction Use of Text Features:

Retelling Sequence of Events (Level 4-24): "Important Events" means the story structure elements or text structure elements.
- Story Structure events are:
  - Beginning: Status Quo (what it's like for the main character and setting before the problem)
  - Middle: Problem (what the main character wants or wants to fix)
    - Try/Fail Cycles or Blocks (instances that attempt to solve the problem but don't)
  - End: Resolution (how the problem is fixed)
    - Ending (what happened after problem was solved)

To reach a 3 on the rubric elements from Beginning, elements from Middle and elements from the End needs to be in the summary.

- Non-Fiction Text Structure is:
  - Title
    - I. Heading
      - A. Important Idea (Main Idea)
        - 1...Details, examples, definitions, background knowledge

Scaffolded Summary (Level 28-80): Same as above, except few (!) details should be included. Summary is mostly Important Ideas.

Retelling: Characters and Details (Level 4-24): If they have mispronounced a name wrong while reading it doesn't count against them to mispronounce it in the oral retelling.

Vocabulary (Level 4-38): We want students to use their own words in retellings and summaries, so using all the vocabulary in the text is not a standard we are encouraging. The old rubric stated the teacher should assess limited/some/basic/good understanding without giving a definition of how to determine this.

We still value learning and using vocabulary from reading. The compromise is that if they use 1 important word score it a 1, if they use 2 then score it a 2, 3 then score it a 3, 4 or more give a 4.

Literal Comprehension (Level 28-80): The text said that, answer is not implied. "Accurate" means copy from the book. No interpretation involved.

Interpretation (Level 18-80): Inference that requires students to extend or combine ideas to explain what the reader thinks, or to explain how or why something happened.
- **Important** Text Implications: Interpretation inference is directly related to the relevant story structure elements.
- **Insightful:** "Wow" factor—I would not expect a child this age to say that.

Reflection (Level 4-80): Inference looking back into the story to say what the big ideas are in the story
- **Significant Event:** Specific event is stated that directly connects to relevant story structure (Beginning, Problem, Try/Fail, Resolution, Ending) or text structure.
- **Relevant Reason:** Reason(s) given connects the significant event to the inference made

Connection (Level 4-16): "Thoughtful" means additional inferential explanation is given beyond connecting to understanding the story/text.