2014

Effect of an Intensive Intervention on the Early Literacy Skills of Kindergarten Students Identified as Most at Risk for Future Reading Difficulties

Craig S. Buscher
The University of Montana

Effect of an Intensive Intervention on the Early Literacy Skills of Kindergarten Students Identified as Most at Risk for Future Reading Difficulties

Craig S. Buscher
Master of Arts in Teaching, Miami University, 1999

Submitted in partial fulfillment for the degree of

Doctorate of Education in Curriculum and Instruction
The Phyllis J. Washington College of Education and Human Sciences
University of Montana
Summer, 2014

Dr. Sandy Ross, Graduate School, Dean

Dissertation Committee:

Dr. Trent Atkins, Chair
Department of Curriculum & Instruction

Dr. Morgen Alwell
Department of Curriculum & Instruction

Dr. Matthew Schertz
Department of Curriculum & Instruction

Dr. Lucy Hart Paulson
Department of Communicative Sciences and Disorders

Dr. Anisa Goforth
Department of Psychology

Abstract

Effect of an Intensive Intervention on the Early Literacy Skills of Kindergarten Students Identified as Most at Risk for Future Reading Difficulties

Craig S. Buscher

Chair of the Committee:
Dr. Trent Atkins
The Phyllis J. Washington College of Education and Human Sciences
University of Montana

There is limited research investigating the individualized effects of early literacy interventions on kindergarten students who are most at risk for reading failure at school entry. In this study, a multiple baseline design was used to measure the effect of an early literacy intervention on the alphabet knowledge and phonological awareness of kindergarten students identified as most at risk. First, a two-stage screening process consisting of measures of alphabet knowledge, phonological awareness, and rapid automatic naming was employed to identify those students with the lowest overall skill level from the entire at-risk kindergarten population from one school building. The nine students with the lowest skill level were selected to receive more intensive intervention services than were typically offered to at-risk kindergarten students. The intervention consisted of both code-focused and meaning-focused components and was delivered daily for 25 minutes in small groups of three students for a total of 60 instructional sessions. The intervention was implemented at three different points in time, resulting in three baseline and three intervention phases. During baseline and intervention phases, two alphabet knowledge measures and one phonological awareness measure were repeatedly administered to all participants. Data was analyzed through systematic comparison of within and between phase patterns, such as performance level, trend, variability, non-overlap of data points, and immediacy

of effect. Analysis was enhanced using the conservative dual criterion approach. Results indicated that an experimental effect was evident after the first and second introduction of the independent variable but not at time three, weakening the claim of a cause and effect relationship between the independent and dependent variables. Differences in performance level and trend in the data were evident for five students on letter name knowledge, six students on letter sound knowledge, and six students on first sound identification skills.

Acknowledgements

I would first like to acknowledge my dissertation committee for their time, guidance, and support through this process. Your insightful feedback and collective wisdom were instrumental in refining this dissertation.

A special thank you to Dr. Morgen Alwell for your guidance pertaining to methodological details and counsel during the planning and implementation stages of this study. Thank you also for expanding my knowledge base in a great number of topics. My sincerest gratitude to Dr. Lucy Hart Paulson for our numerous dialogues related to early literacy instruction, assessment, and research. These exchanges and your guidance were truly invaluable and greatly enriched my experience as a doctoral student. A special thank you to Dr. Trent Atkins for encouraging and supporting me throughout my doctoral experience. I am inspired by your dedication to improving the learning outcomes of all students and your creative ideas for making this happen. I greatly appreciate all of your feedback and guidance as I navigated through this process.

Thank you to the teachers and staff of the school site where this study took place. I am most grateful for your willingness to collaborate on this project. Your advice, suggestions, flexibility, and support were greatly appreciated.

I would like to acknowledge Dr. Tom Kratochwill and Dr. Chris Swoboda for your correspondences related to the Conservative Dual Criterion Approach. Thank you for providing me with resources to enhance my data analysis.

Thank you to my family and friends who supported and encouraged me as I made the plunge back to full-time student. Last, I wish to thank my wife, Sue Tirrell. Your unconditional

support and all of the hard work and sacrifices you made during this journey mean more to me than words can express. Thank you for believing in me.

Table of Contents

List of Tables

List of Figures

CHAPTER ONE: INTRODUCTION TO THE STUDY

Context of the Study

In 2001, Congress passed the No Child Left Behind (NCLB) Act, a reauthorization of the Elementary and Secondary Education Act. NCLB includes Reading First, an initiative providing additional funding for scientifically-based reading programs at kindergarten to third-grade levels in an effort to improve reading outcomes (Lyon, Shaywitz, Shaywitz, & Chhabra, 2005). NCLB further places an emphasis on evidence-based practices (EBP) and accountability in K-12 education (Salvia, Ysseldyke, & Bolt, 2007). The most recent reauthorization of the Individuals with Disabilities Education Act (IDEIA, 2004) is aligned with NCLB in promoting accountability for results for students with academic difficulties, recognizing and attempting to enhance the role of parents in the educational process, and aiming to improve student achievement through evidenced-based practice and access to high-quality curricula and instruction (Salvia et al., 2007). Ikeda (2012) noted that IDEIA (2004) and NCLB collectively focus on the performance of students with disabilities in the general education curriculum. IDEIA further includes language allowing schools to implement evidence-based programs in a multi-tiered instructional framework commonly referred to as Response to Intervention (RTI). Fifteen percent of funds from IDEIA can now be used by schools for early intervention services for all students.

In addition to these legislative acts, the National Research Council (NRC) (Snow, Burns, & Griffin, 1998), the National Reading Panel (NRP) (2000), and the National Early Literacy Panel (NELP) (2008) have all published reports within the last sixteen years pertaining to empirical findings of early literacy and reading research. Consequently, recent instructional

practices, particularly in the case of early reading instruction, have been increasingly influenced by the convergence of scientific evidence (Lyon et al., 2005). Lyon et al. (2005) noted that reading has become the focal point of scientifically-based instruction over other academic areas, in part because reading proficiency is so fundamental to academic success and quality of life. Reading interventions delivered in the context of a multi-tiered framework have also become more prevalent (Gersten et al., 2009). The current widespread use of RTI is evident from a recent survey from a nationally representative sample of school districts, in which Bradley et al. (2011) found that approximately 71% of school districts surveyed were using RTI.

Despite this emphasis on evidence-based practices and focus on performance, significant numbers of students in America's schools continue to read at or below basic levels of reading proficiency. In 2011, approximately 33% of fourth graders and 24% of eighth graders scored below a basic level in reading and only 34% of students in both grade levels scored at or above proficiency as measured by the National Assessment of Educational Progress (NAEP) (http://nationsreportcard.gov/reading_2011/summary.aspx). Converging evidence suggests that effective early reading intervention in kindergarten and first grade can bring most students who are at below an average level of reading performance to an average level (Al Otaiba & Fuchs, 2006; Al Otaiba et al., 2011; Cavanaugh, Kim, Wanzek, & Vaughn, 2004; Scanlon, Vellutino, Small, Fanuele, & Sweeney, 2005; Torgesen et al., 2001; Vellutino et al., 1996; Vellutino, Scanlon, Small, & Fanuele, 2006). For example, in a synthesis of 27 quantitative studies involving school-based reading interventions delivered to kindergarten students, Cavanaugh et al. (2004) concluded that interventions are effective in improving early literacy skills of students identified as at risk for future reading difficulties. These studies support the premise of early intervention in reading.

Simmons et al. (2013) refer to kindergarten "as a critical window of opportunity in which intervention can differentially accelerate reading growth compared to later intervention for children at early reading risk" (p. 2). Through early identification and intervention practices, schools can deliver reading supports to kindergarten students early in the school year as opposed to delaying support services toward the second half of kindergarten or beyond. How to identify students as at risk and what level of supports to deliver are important questions that still need to be addressed (Vellutino, Scanlon, Zhang, & Schatschneider, 2008). In the remainder of this chapter, I will first discuss broad findings related to effective early literacy instruction. Next, I will provide an overview of the RTI framework, followed by a specific focus on instructional intensity and delivery of supports in a three-tiered RTI model. Finally, characteristics and predictors of students who do not respond to early reading intervention will be discussed.

Early Intervention in Reading

Early literacy instruction consists of a number of components that can be characterized as either code focused or meaning focused in nature. Code-focused instruction consists of components such as alphabet knowledge, phonemic awareness, word identification, decoding, and spelling. Meaning-focused instruction consists of components such as oral language skills, listening comprehension skills, vocabulary development, and story structure (Al Otaiba et al., 2011). Taken together, reports from the National Reading Council, the National Reading Panel (2000), and the National Early Literacy Panel (2008) have highlighted major research findings pertaining to the prevention of reading difficulties, key code-focused and meaning-focused components of early literacy instruction, and how best to teach these components.

According to the National Research Council report (Snow et al., 1998), word recognition skills and knowledge of text content make separable contributions to reading comprehension. To

become proficient at word recognition, children must develop a strong understanding of the alphabetic principle. In turn, one important aspect of understanding the alphabetic principle is developing phonological awareness (PA), or awareness that spoken language can be analyzed into strings of separable words and words into sequences of syllables and individual phonemes, or speech sounds. When children receive sufficient instruction in phonemic awareness, reading and spelling growth are accelerated. On the other hand, when children have difficulty with phonemic awareness, difficulties arise in sounding and blending new words, retaining words from one encounter to the next, and in learning to spell. Both knowledge of letter symbols and sounds and a working awareness of the phonemic composition of words are necessary for proficient decoding and spelling (Snow, Burns, & Griffin, 1998).

The National Reading Panel (2000) reported findings pertaining to the "how" of reading instruction in the areas of phonemic awareness, phonics, fluency, vocabulary, and comprehension. Major findings related to kindergarten instruction will be briefly summarized. First, the NRP found that explicit instruction in phonemic awareness led to improvements in reading and spelling. The most effective instruction consisted of factors such as manipulating phonemes with letters, focusing on fewer types of phoneme manipulations, and teaching students in small groups. Second, the NRP concluded that kindergarten students who received systematic beginning phonics instruction improved in their ability to read and spell. The panel found positive and significant effects for students with disabilities, as well as for students with low achievement and from a low socioeconomic background. Third, the NRP found that vocabulary can be learned incidentally through storybook reading and in listening to others. Repeated exposure to words and encountering words in a number of contexts appeared to aid in vocabulary acquisition.

Following their review of a large number of studies involving preschool and kindergarten students, the National Early Literacy Panel (2008) reported the following six skills as having consistently moderate to strong relationships to the later literacy outcomes of decoding, reading comprehension, and spelling: 1) alphabet knowledge (AK), 2) phonological awareness, 3) writing or name writing, 4) rapid automatic naming of objects and colors, 5) rapid naming of letters and digits, and 6) phonological short-term memory. Oral language skills had a moderate relationship to later literacy outcomes, but were one of the weakest predictors in this range. The average correlation of oral language to reading comprehension increased however when outcome measures were in first or second grade as opposed to kindergarten.

Early literacy interventions reviewed by the NELP (2008) were categorized as either having a code focus or a shared-reading/sharing books focus. The NELP characterized the effect sizes (ESs) reported below as small up to 0.30, moderate if in the 0.50-0.79 range, and large if 0.80 or higher. The NELP analysis of code-focused interventions produced the highest ES for phonological awareness (0.82). The effect sizes for reading, writing, and spelling were 0.44, 0.61, and 0.61 respectively. Impacts for print knowledge, alphabetic knowledge, and oral language were 0.47, 0.38. and 0.32 respectively. The NELP found the largest effect sizes on reading outcomes and significantly higher effect sizes for AK when the code-focused interventions combined phonological awareness training with phonics. Interestingly, only when both the phonemic tasks of segmenting and blending were combined in the intervention were ESs for both reading and spelling outcomes noteworthy (0.49 and 0.56 respectively). In summary, code-focused interventions had a large to moderate effect on predictors of later reading and writing, such as PA and AK, and on reading and writing outcomes of both preschool and kindergarten students.

Shared-reading interventions had large effect sizes for oral language (0.57 with one outlier removed) and print knowledge (0.50). Few shared-reading studies included outcome measures for AK, PA, reading readiness, reading comprehension, decoding, and spelling; therefore making it difficult to determine reliable effect sizes. Effect sizes for measures of simple vocabulary and composite oral language measures were 0.60 and 0.35 respectively. Shared-reading interventions that were more interactive in nature and more intensive in terms of frequency had a more positive impact on print knowledge and oral language. Although it is clear shared-reading interventions improve oral language skills across variations in age and risk factor, evidence does not exist on their effectiveness on other emergent literacy skills or later literacy outcome measures. Consequently, the NELP (2008) concluded that shared-reading interventions should be used in conjunction with code-focused interventions.

Collectively, findings from the NRC (1998), the NRP (2000), and the NELP (2008) make clear the importance of both code-focused and meaning-focused components of early literacy instruction. In addition to instruction in phonemic awareness, alphabet knowledge, and phonics, results suggest the importance of integrating these components. Additionally, including both code- and meaning-focused components in an early literacy intervention appears to be necessary. In the next section, I turn my attention to instructional delivery within an RTI framework.

Response to Intervention

The RTI framework is comprised of a number of core principles. Barnes and Harlacher (2008) identified the core principles as follows: a preventive and proactive approach to education, matching instruction to student needs, data-based decision making, a problem-solving orientation, use of evidence-based practices, and a systems level approach. In addition, RTI models utilize a more formative and diagnostic approach to assessment that allows educators to

continuously evaluate instructional effectiveness and modify instruction based on identified areas of student need. Based on these assessment results, students who are not responding adequately to instruction receive increasing levels of instructional intensity.

Typically, RTI models consist of three tiers of instructional intensity. The primary tier of instruction, or Tier One, refers to the core instruction received by all students in the general education setting (Gersten et al., 2009). At Tier One, all students are assessed through universal screening measures to determine their level of academic performance in areas such as reading and math. Students who are below grade-level expectations, or who are deemed nonresponsive to Tier One supports as indicated by progress monitoring assessment tools, receive supplemental supports. These Tier Two supports are delivered in small groups and are intended to be in addition to, not in place of, Tier One instruction. Students who continue to be nonresponsive to instruction at Tiers One and Two receive even more intensive instruction targeted to their individualized needs, referred to as Tier Three supports (Gersten et al., 2009).

Regardless of the level of support, the RTI model prescribes the use of evidence-based instruction in an effort to elevate the greatest number of students to a level of academic proficiency. This begins with effective reading instruction at the Tier One level that enables about 80% of the school-wide population to reach levels of reading proficiency (Al Otaiba et al., 2011). Additionally, core classroom (Tier One) instruction is important in enhancing the learning of students at risk for reading failure (Slavin, Lake, Davis, & Madden, 2011), as well as reducing the number of students at risk and promoting positive outcomes for students identified with a learning disability (LD) (Fuchs & Vaughn, 2012). In order to sustain the gains of students who have received intervention services, Slavin et al. (2011) further stressed the necessity of effective classroom models at Tier One.

Intensity of Instruction

In Tiers Two and Three, evidence-based instruction continues and intensity of instruction increases. Intervention intensity can be increased in a number of respects. According to Vaughn, Denton, and Fletcher (2010), interventions can be intensified by increasing the frequency (number of sessions per week), the duration (number of weeks and months the intervention is provided), or the length of sessions. Interventions can also be delivered in multiple sessions per day.

Another way to intensify instruction is by varying the nature of the intervention (Vaughn et al., 2010). For example, instruction can be made more systematic and explicit with increased teacher scaffolding and support, immediate corrective feedback, and more opportunities for students to respond and practice skills in isolation. Additionally, Fuchs and Fuchs (2006) considered teacher expertise as a component of instructional intensity. Although Fuchs and Fuchs did not define teacher expertise, Slavin et al. (2011) identified instruction delivered by certified teachers and reading specialists as one component of the gold standard among intervention delivery. In their synthesis, Slavin et al. found that teachers were more effective tutors than paraprofessionals and volunteers. They concluded that schools can use a mix of teachers, paraprofessionals, and volunteers to deliver intervention services but should assign certified teachers to work with children who have the greatest needs. Compton et al. (2012) reported that prior research suggests that Tier Three intervention should not be viewed as a more intensive Tier Two with the same techniques and materials. Instead, Tier Three intervention should be of appropriate intensity and with sufficient individualization to meet students' needs.

Instruction can also be intensified by adjusting group size. Although a number of recommendations exist, consensus of ideal group sizes for Tier Two and Three reading

interventions in the primary grades has not yet been reached. Small groups are generally considered to be adequate for Tier Two interventions, although some contradictory evidence exists. Gersten et al. (2009) recommended groups of four to five students with homogeneous needs, while Vaughn et al. (2010) suggested group sizes between one-to-one and one-to-three. The Gersten et al. recommendation followed a review of eleven studies pertaining to Tier Two interventions, in which no significant effect-size differences were found between instruction delivered in small groups as compared to one-to-one. Since effect sizes were similar, the authors contended that delivering Tier Two interventions in small groups would be the more practical alternative in school settings. In contrast, Slavin et al. (2011) suggested the use of one-to-one interventions at Tiers Two and Three following their synthesis of reading interventions for students aged five through ten. Fuchs and Fuchs (2006), in their discussion of instructional intensity, noted that intervention groups are not just smaller but more homogeneous. Based on these reviews and recommendations, further study is needed to investigate ideal group sizes for students with the most intensive needs.

Delivery of Supports in a Three-Tiered Model

Within an RTI process, school personnel make decisions regarding which students need and do not need additional supports based on assessment data. In a three-tiered model in which Tier Three is not synonymous with a special education evaluation, the implication is that some students will benefit sufficiently from Tier One instruction, others will need additional supplemental supports to be successful, others will require intensive supports to be successful, and yet others will need long-term supports offered through special education services. In this model, two possibilities exist for the delivery of support services. The first possibility is that all students in need initially receive Tier Two levels of support. The second possibility is that some

students receive Tier Two levels of support whereas others immediately receive Tier Three levels of support. In the first scenario, some students will receive intervention supports at a level insufficient to move them to a level of academic proficiency, and they will inevitably receive more intensive supports. In the second scenario, it becomes somewhat necessary to be able to reliably predict which students require intensive supports to be successful. Ikeda (2012) commented that schools need ways to know when to put more resources toward students that need them.

With students in upper elementary and middle school, Fuchs, Fuchs, and Compton (2010) argued that older students with sizable academic deficits should receive immediate Tier Three supports. Vaughn et al. (2010) noted that older students with severe reading difficulties require highly intensive interventions over several years. They further suggested moving students with low reading scores in grades three and higher directly into Tier Three. However, some evidence exists to support the assertion that more intensive, Tier Three supports should be provided more quickly for students in the early elementary grades as well. Following an analysis of five early reading intervention studies, Torgesen (2000) concluded that researchers have yet to discover the conditions necessary for children with the most severe difficulties in the early elementary grades to acquire adequate reading skills. In a reading intervention study by Al Otaiba and Fuchs (2006) that spanned kindergarten and first grade, 92% of students who were nonresponsive after kindergarten were also nonresponsive after first grade. Students received interventions between 5-20 minutes three times per week. O'Connor (2000) also conducted a two-year reading intervention study beginning in kindergarten and ending at the end of first grade. Interventions were delivered in four layers of increasing intensity to students designated as highly at risk. Although students with disabilities who participated in all four layers of the intervention

outperformed those who did not participate on measures of word reading, nine of these twelve students were still characterized as poor readers at the end of first grade. Findings from all three of these studies suggest the need for more intensive reading supports for some students starting in kindergarten.

Nonresponsiveness to Intervention

Although instruction in an RTI model is delivered in increasing levels of support based on student need, this does not guarantee that all students will respond to instruction. Previous studies have attempted to identify early predictors or characteristics of nonresponders to early reading interventions (Al Otaiba & Fuchs, 2002; Al Otaiba & Fuchs, 2006; Catts, Fey, Zhang, & Thomblin, 2001; Fuchs et al., 2012). Understanding these characteristics will aid in the early identification of students who will be unlikely to respond to Tier One and Two supports. Students who have been identified in the at-risk category through the universal screening process can be given a more thorough assessment battery that measures characteristics of nonresponders. These students can then be differentiated to receive more or less intensive intervention supports.

At present, although research on the effectiveness of reading interventions is more prevalent in the early elementary grades than in upper elementary and middle school levels (Wanzek, Wexler, Vaughn, & Ciullo, 2010; Vaughn et al., 2008), considerably less is known about effective interventions for students at all grade levels who did not respond to Tiers One and Two (Vaughn & Fuchs, 2006) and require interventions that are both extensive in duration and intensive in nature (Chard, 2012; Wanzek & Vaughn, 2007). Wanzek & Vaughn (2007) conducted a synthesis of extensive, early reading interventions, defined as one hundred sessions or longer, or 20 weeks of daily instruction. The results of Wanzek and Vaughn's synthesis indicated that reading interventions beginning in first grade "are associated with higher effects

than interventions beginning in second or third grade" (p. 557). They further noted that the reading difficulties of students in second and third grade are more complex, making it more difficult to achieve substantial reading gains. Although Chard (2012) pointed out that students in an RTI model receive immediate instructional support based on nonresponsiveness to core instruction, Vaughn, Denton, and Fletcher (2010) noted that Tier Three supports may be necessary to avoid another wait-to-fail model if Tier Two interventions are provided at low levels of intensity. For students in grade one, Vaughn et al. recommended preventive Tier Two intervention delivered in groups of three to four students, four to five days per week, for 20-30 weeks but made no recommendations for the most at-risk students in kindergarten. Theoretically, intensive intervention supports can be implemented immediately if these students can be accurately identified in the beginning of their kindergarten school year.

To summarize Chapter One, NCLB (2001) and IDEIA (2004) together promote accountability for results for students with academic difficulties, aim to improve student achievement through evidenced-based practice and access to high-quality curricula and instruction (Salvia et al., 2007), and emphasize student performance (Ikeda, 2012). National reports from the NRC, NRP and NELP inform our current understanding of early literacy teaching and learning. Response to Intervention provides a preventive and proactive framework from which to deliver interventions in increasing levels of instructional intensity based on assessment data collected in a consistent and frequent manner. Research elucidating reliable predictors of response to intervention can inform appropriate levels of instructional intensity to elevate the greatest number of students to proficient levels of reading.

Statement of the Problem

A review of the literature has identified a lack of research pertaining to reading interventions for those kindergarten students with the most intensive needs. Some researchers advocate for the more immediate use of intensive, Tier Three supports for students with the most severe reading difficulties. However, this recommendation has been limited to students in grades three and beyond. In kindergarten through second grade, students receive a continuum of supports starting from Tier One and progressing to Tier Two before more intensive Tier Three supports are made available. Evidence suggests that some of these students are unlikely to respond to Tier One and Tier Two instruction even if well-designed, evidence-based programming is in place (see Torgesen, 2000). As such, requiring a progression of supports to all students in the early elementary grades, regardless of their current level of performance, will likely result in an insufficient path to reading proficiency for a subgroup of the population. Recent research has identified more reliable methods for predicting which students in kindergarten are most at risk and less likely to respond adequately to Tier One and Tier Two instruction (Vellutino et al., 2008). A service delivery model in which intensive supports are delivered more immediately to these students is needed.

Purpose of the Study

The purpose of this single-case, multiple baseline across participants study was to investigate the effectiveness of an early literacy intervention on the alphabet and phonemic awareness skills of kindergarten students with the most intensive early literacy needs. Effectiveness was determined by repeated measurement of letter-name knowledge, letter-sound knowledge, and first sound identification during baseline and intervention phases.

Research Questions

Consistent with the purpose of the study, I posed the following three questions: 1) Is an intensive early literacy intervention effective in improving the performance level and rate of improvement of letter name knowledge of most at-risk kindergarten students beyond baseline levels? 2) Is an intensive early literacy intervention effective in improving the performance level and rate of improvement of letter sound knowledge of most at-risk kindergarten students beyond baseline levels? 3) Is an intensive early literacy intervention effective in improving the performance level and rate of improvement of phonemic awareness, as measured by first sound identification, of most at-risk kindergarten students beyond baseline levels?

Perspective of the Researcher

I worked as a public school teacher for nine years, eight of which as a special education teacher or reading specialist. The vast majority of this time I implemented reading interventions to students with learning disabilities and students identified with reading difficulties within an RTI framework. In my first year of this work, I began receiving what I consider to be quality professional development in language structure, literacy instruction, literacy assessment, and intensive literacy interventions. During this time I came to several realizations: 1) there was much I did not know about teaching students how to read, 2) the overwhelming majority of students with reading difficulties could make significant and consistent progress on the road to proficient reading with proper instruction, and 3) I had a desire to train other teachers.

I continued to receive professional development over the next several years and attend trainer of trainer sessions in order to return to my district and teach others. The school in which I worked was chosen to be a pilot school for RTI for the state. I received RTI consultant training to assist area schools with their own RTI implementation. My professional experiences have

taught me the importance and value of early and continued intervention, high expectations, linking assessment to teaching, and explicit, systematic instruction. These experiences were instrumental in leading me to the area of research I am exploring in this study and in my current role as a pre-service teacher educator.

Definition of Terms

alphabet knowledge. Knowledge of the names and sounds associated with printed letters (NELP, 2008).

code-focused instruction. Instruction including components such as alphabet knowledge, phonemic awareness, word identification, decoding, and spelling (Al Otaiba et al., 2011).

decoding. Ability to translate a word from print to speech through knowledge of sound-symbol correspondences (Moats, 2010).

intensive early literacy intervention. In this study, instruction delivered daily in groups of three students for 25 minutes a day for a total of 60 instructional sessions.

Meaning-focused instruction. Instruction including components such as oral language skills, listening comprehension skills, vocabulary development, and story structure (Al Otaiba et al., 2011).

nonresponders. Students who do not respond adequately to Tier One and Tier Two instruction.

oral language. The ability to produce or comprehend spoken language, including vocabulary and grammar (NELP, 2008).

phoneme. An individual speech sound (Moats, 2010).

phonemic awareness. The conscious awareness that words consist of smaller speech segments, or individual speech sounds (Moats, 2010).

phonics. The study of relationships between letters and the sounds they represent (Moats, 2010).

phonological awareness. Awareness and sensitivity to the different units of spoken language at the word, syllable, and phoneme level.

progress monitoring. Repeated measurement of academic performance to evaluate the effectiveness of instruction (Center on Response to Intervention, http://www.rti4success.org/essential-components-rti/progress-monitoring).

rapid automatic naming (RAN). The ability to rapidly name a series of items presented visually in a random, repeated fashion. Items typically pertain to one category, such as letters, digits, objects, or colors and represent only a small subset of that category (NELP, 2008).

Response to Intervention (RTI). A schoolwide, systems approach to instruction and assessment that includes but is not limited to evidence-based instruction, data-based decision making, matching instruction to student need, and increasing levels of instructional intensity (Barnes & Harlacher, 2008).

single subject, multiple-baseline across participants design. An experimental design to document the causal relationship between independent and dependent variables in which the independent variable is introduced at staggered times with three or more data series (Horner et al., 2005).

Tier One instruction. Instruction delivered to all students in a particular grade or classroom that consists of high-quality core curricula and research-based instructional practices (Center on Response to Intervention, http://www.rti4success.org/essential-components-rti /progress-monitoring).

Tier Two instruction. Evidence-based instruction delivered in small groups to students who have not adequately responded to Tier One instruction as indicated through data collection and analysis pertaining to pre-established levels of performance.

Tier Three instruction. Evidence-based instruction delivered one-to-one or in small groups to students who have not adequately responded to Tier One and Tier Two instruction as indicated through data collection and analysis pertaining to pre-established levels of performance.

universal screening. An assessment process in which all students are given brief assessments in skill areas such as literacy and math, typically three times per school year.

CHAPTER TWO: REVIEW OF THE LITERATURE

A review of the literature was conducted of early literacy intervention research for kindergarten students. The conceptual framework for this study is the convergence of two main bodies of literature. The first body pertains to quantitative studies investigating the instructional effectiveness of early literacy intervention for kindergarten students at risk for future reading failure. Specifically, kindergarten reading intervention research delivered in a multi-tiered or layered instructional framework was targeted. As this research is still emerging, studies of intensive kindergarten interventions that preceded the current emphasis on multi-tiered frameworks were also reviewed. For the purposes of this review, these studies have been categorized into three areas: a) less intensive kindergarten intervention studies, b) more intensive kindergarten studies, and c) studies that compared less and more intensive interventions. I categorized studies as less or more intensive based on the following factors: duration of the intervention, length of intervention sessions, frequency of sessions, and group size. If the study continued beyond kindergarten, the distinction between less and more intensive intervention was made based on specifics of the kindergarten intervention only. Most studies reviewed in the above three categories are described in detail, with particular attention to both features of instructional intensity and major findings. Throughout each section, study aspects are connected and synthesized.

The intervention studies in this first body of literature typically consisted of multiple components of early literacy instruction. These components can be characterized as either code-focused or meaning-focused in nature. Code-focused instruction consists of components such as alphabet knowledge, phonemic awareness, word identification, decoding, and spelling. Meaning-focused instruction consists of components such as oral language skills, listening

comprehension skills, vocabulary development, and story structure (Al Otaiba et al., 2011). The researcher of this study is primarily focused on code- and meaning-focused components especially relevant to beginning kindergarten instruction. As many of these multi-component intervention studies either began in the latter half of kindergarten or lacked specifics regarding the components, organization, and sequencing of alphabet knowledge instruction, studies that informed and added to the literature base in this domain were also examined. In addition, studies that included meaning-focused components tended to consist of activities revolving around shared-book reading. As many of these studies again lacked specifics pertaining to the components and organization of meaning-focused instruction, literature on shared-book reading interventions pertaining to kindergarten students was reviewed.

The second main body of research pertains to studies investigating predictors or characteristics of intervention responsiveness for kindergarten students. Typically, researchers of these studies collected assessment data prior to instruction on a number of constructs, such as phonological awareness, rapid automatic naming, alphabet knowledge, behavior, verbal abilities, and vocabulary. These same measures were administered again throughout or at the end of the study. Because kindergarten students often begin the school year as prereaders, additional reading outcome measures of word reading and reading comprehension were often included at posttesting only. Researchers then conducted statistical analyses to identify those variables that best predicted performance on later reading outcome measures. Knowledge of the best predictors of future outcomes can be of value in early identification of students at risk for future reading difficulties, and in turn lead to a more proactive and preventive approach to early literacy instruction.

The review of the literature has been organized as follows. Studies of less intensive kindergarten studies are discussed and analyzed first, followed by studies that involved more intensive instruction pertaining to features such as frequency, duration, and group size. Next, studies that compared less and more intensive intervention conditions are discussed and analyzed. Studies investigating the components and organization of alphabet knowledge instruction and shared-book reading follow. Finally, studies investigating predictors of intervention responsiveness for kindergarten students are discussed and analyzed.

Less Intensive Kindergarten Interventions

A number of less intensive, early literacy intervention studies have been conducted for kindergarten students who were identified as at-risk for future reading failure (Fuchs et al., 2002 (unpublished), as described in and further cited as Al Otaiba & Fuchs, 2006; Gyovai, Cartledge, Kourea, Yurick, & Gibson, 2009; Musti-Rao & Cartledge, 2007; O'Connor, 2000; Samanich, 2003; Scanlon, Vellutino, Small, Fanuele, and Sweeney, 2005). For studies that continued beyond kindergarten, outcomes in subsequent years are also discussed.

All studies reviewed were experimental in nature. Two of the studies reviewed utilized pretest-posttest comparison group designs (Al Otaiba & Fuchs, 2006; Scanlon et al., 2005). Some studies utilized single subject experimental designs, specifically multiple baseline across participants designs (Gyovai et al., 2009; Musti-Rao & Cartledge, 2007; Samanich, 2003). The design used in O'Connor (2000) was not clearly defined. Pretest-posttest group designs are discussed first, including the study by O'Connor, followed by a discussion of the multiple baseline studies.

All of the pretest-posttest studies were conducted during kindergarten and first grade (Al Otaiba & Fuchs, 2006; O'Connor, 2000; Scanlon et al., 2005). Inclusionary criteria for participant selection were not discussed in Al Otaiba & Fuchs (2006). The two other studies used a different process and criteria for identifying students at risk. However, a letter naming or letter identification task was used in both studies, and O'Connor (2000) also used a measure of phonological awareness. Scanlon et al. (2005) defined at-risk status as performance below the 30th percentile on the Letter Identification (LID) subtest of the Woodcock Reading Mastery Test-Revised (WRMT-R; Woodcock, 1987). O'Connor's (2000) initial criteria for at-risk status consisted of a combined score of less than an 86 standard score (SS) on the Letter-Word Identification and Dictation subtests of the Woodcock-Johnson (WJ) Tests of Achievement (Woodcock & Johnson, 1990), less than four segments on a phonemic segmentation task, and less than 15 letter names on a letter naming task in October of kindergarten.

Two of these studies consisted of multiple intervention conditions (Al Otaiba & Fuchs, 2006; Scanlon et al., 2005). In the kindergarten portion of their study, Scanlon et al. randomly assigned participants to the intervention condition (n = 232) or the comparison condition (n = 230). Intervention began in early October and ended in early June. Sessions occurred twice per week for 30 minutes in small groups of three students. The kindergarten intervention targeted early literacy skills that tend to be lacking in students with early reading difficulties, such as phonemic awareness, letter-name and letter-sound knowledge, an understanding of the alphabetic principle, concepts of print, and whole word identification. In first grade, participants were assigned to one of two experimental conditions, a phonological skills emphasis (PSE) condition and a text emphasis (TE) condition, or the comparison condition. The PSE condition involved 15 minutes of instruction in phonological skills and five minutes each for reading and rereading

of text, sight word instruction, and writing. The TE condition involved 15 minutes of instruction in reading and rereading of connected text and five minutes each for phonological skills, sight word instruction, and writing. Students in the comparison conditions in both grades received school-based interventions. Both conditions consisted of one-to-one instruction for 30 minutes on a daily basis.

Al Otaiba & Fuchs (2006) randomly assigned 312 students in kindergarten to one of three conditions. One group of students received teacher-directed phonological awareness activities from O'Connor, Notari-Syverson, and Vadasy's (1998) Ladders to Literacy intervention program for five-fifteen minutes, three times per week for 20 weeks. The second group received a combination of the Ladders to Literacy activities and instruction from the Kindergarten Peer-Assisted Learning Strategies (K-PALS; Mathes, Clancy-Menchetti, & Torgesen, 2004) literacy intervention program, which was delivered in 20-minute sessions, three times per week for 16 weeks. The size of the groups was not mentioned. The third group was the control condition which received typical classroom instruction.

In first grade, students who participated in the kindergarten intervention either received first-grade intervention or typical classroom instruction. Additionally, some students who did not participate in the kindergarten intervention also received intervention services. The first grade intervention consisted of the First-Grade PALS program delivered for 20 minutes, three times per week for 20 weeks.

The study by O'Connor (2000) was organized somewhat differently. Interventions were delivered in 4 layers of increasing intensity during certain time periods of kindergarten and first grade to students designated as high at risk. Layer 1 was delivered in a whole group setting and

included activities from Ladders to Literacy and focused on teaching letter names and sounds, attending to the sounds in spoken words, and speaking and listening skills. Children participated in three or more sessions a week from October to January. Layer 2 began in February and was supplemental to Layer 1 instruction. Layer 2 consisted of one-to-one instruction delivered in 12 min. sessions, three times per week over a 10-week period. Students were taught letter names and sounds, phonemic blending and segmentation activities, and matching letter sounds to words. Of the 44 students who met inclusion criteria from the January testing cycle, teacher willingness to participate and parental consent was obtained for only 25 students.

Layer 3 instruction consisted of small groups of three to five children for 30 minutes, four times per week over a 14-week period. However, five of these students also received 20-30 minutes of reading practice, two to three times per week with a special educator. Instruction began in November of first grade and focused on letter sounds, phonemic decoding, spelling, and reading of controlled texts. Of the 20 students who still met the inclusion criterion, 14 made gains above the average gain of typical readers on phonological blending and segmenting outcome measures. However, four of these students were still characterized as high at-risk since their beginning level of performance in October of first grade on the combined subtests of the WJ was still below 86. These four students, the six who did not make adequate gains in Layer 3, and two students with Individualized Education Programs (IEPs) received instruction in Layer 4. Layer 4 consisted of one-to-one instruction in blending and spelling decodable words and was delivered by a researcher for 15 minutes, four times per week over four weeks.

Across all three studies during kindergarten, group sizes ranged from whole group instruction to instruction in small groups to one-to-one instruction. Sessions were held three

times per week in O'Connor (2000) and Al Otaiba & Fuchs (2006), and twice per week in Scanlon et al. (2005). Session durations ranged from twelve to 30 minutes.

Results of the studies varied significantly. Twenty of twenty-five students who participated in Layer 2 in the O'Connor (2000) study still met inclusion criteria for first grade intervention. Perhaps this can in part be explained by lack of intensity in terms of group size, number of sessions, session length, and duration. Results of the kindergarten intervention in Scanlon et al. (2005) are more difficult to interpret. At the beginning of first grade, students who were originally identified as at risk were screened with the Letter Identification, Word Identification (WID), and word Attack (WAT) subtests of the WRMT-R, as well as researcher-developed measures of letter-sound knowledge, sight vocabulary, and alphabetic decoding. Of special interest here is the use of a battery of assessments and their combined scores to determine risk status. Using *z* scores for each of these measures, at-risk status was defined as scoring at or below the midpoint of the summed *z* scores. In effect, half of the children were identified as being at risk. Without a set criterion for response to intervention, the number of students still at risk is undetermined. It would be important to note, however, that sixty percent of the 172 children in the at-risk group came from the kindergarten comparison condition, suggesting that the kindergarten intervention was successful in reducing the percentage of students in need of additional supports.

At the end of first grade, the incidence of poor reading ability was reduced to less than 8% for all three, first-grade groups combined that included students who participated in the kindergarten intervention. However, students in the two, first-grade experimental conditions increased their standard scores at each measurement point and showed more growth relative to national norms than did students in the first-grade comparison condition. No students in the PSE

condition scored below a standard score of 85 on the Basic Skills Cluster of the WRMT-R compared to approximately 8% of students in the other two conditions. A higher percentage of students in the PSE condition also scored in the average range. Students in the TE condition scored higher on the WID than children in the other two groups and students in the PSE condition scored higher on the WAT. Results were less encouraging for students who participated in the first-grade experimental condition but did not receive kindergarten intervention. Approximately 10% and 18 % of students in the PSE and TE conditions respectively scored below 85 on the WRMT-R cluster. However, 25% of students who were in comparison condition during both school years scored below 85. According to Scanlon et al. (2005), results indicated that students who participated in a program that emphasized phonological skills were less likely to be seriously behind their peers in foundational reading skills. Results for reading comprehension as measured at the end of first grade using the Reading Comprehension subtest of the Wechsler Individual Achievement Test (Psychological Corporation, 1992) were more mixed. The only statistically significant difference was found between students in the TE condition and the comparison condition.

Al Otaiba & Fuchs (2006) used a similar method for determining response to intervention. Nonresponsiveness and responsiveness to intervention was determined both at the end of kindergarten and first grade. Nonresponsiveness for kindergarten students was defined as performing in the lowest 30th percentile of intervention students on growth of letter sound and phonemic segmentation fluency measures from pre- to post-assessment. This translated to segmenting less than 13 phonemes (sounds) or identifying less than 12 letter sounds per minute. Responsiveness was defined as scoring at or above the mean of the intervention group's growth of letter-sound and phonemic segmentation. Nonresponsiveness for first-grade students was

defined as reading less than 40 words correct per minute on grade-level text. Responsiveness was defined as at or above the intervention group's mean in oral reading fluency. The authors justified their use of fluency measures from prior research that suggested timed measures were better than untimed measures at categorizing poor readers with and without a learning disability.

At the end of first grade, students were categorized in one of three ways: nonresponsive (not meeting responsiveness criteria in either grade), sometimes responsive (meeting responsiveness criteria in either kindergarten or first grade), and always responsive (meeting responsiveness criteria in both kindergarten and first grade). Overall, 7% of the 227 students who received intervention services for one or two years were deemed nonresponders and 25% of the 71 control students were deemed nonresponders, demonstrating the preventive effect of intervention services.

Three other findings of the original study described by Al Otaiba & Fuchs (2006) are of particular interest. First, almost 92% of the sample and all children with IEPs who were nonresponsive at the end of kindergarten were also nonresponsive after first grade. This finding indicates the need for a more intensive early intervention for these students. Second, only one student who received K-PALS plus Ladders to Literacy was characterized as a nonresponder. Although the reasons for the K-PALS plus Ladders success rate over the Ladders only group cannot be teased out, the more intensive, comprehensive approach was more effective for at-risk kindergarten students in this study. Third, nonresponsive students in the Ladders to Literacy condition were in classrooms with significantly lower quality of instruction ratings than sometimes and always responsive students. Likewise, nonresponsive students in the first-grade PALS condition were in classrooms with significantly lower fidelity than always responsive students. From this study, measuring quality or fidelity of instruction appears to be an important

intervention study component and is important to consider when interpreting results of intervention effectiveness.

Across kindergarten and first grade, all studies reviewed thus far included an increase in instructional intensity. O'Connor (2000) increased instructional intensity during both kindergarten and first grade years. Al Otaiba and Fuchs (2006) and Scanlon et al. (2005) did not increase instructional intensity until the beginning of first grade. In Al Otaiba and Fuchs, instructional intensity was only increased from five to fifteen minutes in kindergarten to 20 minutes in first grade. Scanlon et al. increased instructional intensity more significantly, decreasing group sizes from three to one and increasing session frequency from two days a week to a daily basis. At the conclusion of their article, Scanlon et al. (2005) argued that the structure of the kindergarten intervention made it more feasible to implement. However, 113 children received one-to-one supports for most of their first-grade year. The question remains if a more intensive, small group intervention in kindergarten would have reduced the need for one-on-one supports during first grade.

The other studies characterized as less intensive utilized a multiple baseline across participants design to investigate the effect of early literacy interventions for kindergarten students (Gyovai et al., 2009; Musti-Rao & Cartledge, 2007; Samanich, 2003). With this design, participants are divided into groups and receive intervention with staggered starting times following a baseline condition (Barlow, Nock, & Hersen, 2009). As opposed to comparing results of outcome measures to a comparison group, experimental effect is demonstrated by introducing treatment at different points in time and observing changes in dependent variables (Horner et al., 2005).

Gyovai et al. (2009), Musti-Rao & Cartledge (2007), and Samanich (2003) all implemented the Early Reading Intervention (ERI; Simmons & Kamenui, 2003). Across studies, interventions times ranged from 20 to 30 minutes, two to four days a week, and for seven to sixteen weeks. Group sizes ranged from two to four students per group. All three studies measured students' phonemic segmentation and word reading skills during baseline and intervention phases with the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002) phonemic segmentation fluency (PSF) and nonsense word fluency (NWF) subtests. Within each study, the intervention phases for all groups ended at the same time. Consequently, students received instruction for differing amounts of time. During baseline and intervention phases, researchers collected inter-assessor agreement data for progress monitoring assessments. Researchers also collected treatment integrity data throughout the intervention phase.

Gyovai et al. (2009) implemented the ERI intervention to students who were English language learners (ELL). After two weeks of collecting baseline data, intervention began with the first group. When at least three data points were above baseline levels for students receiving intervention, intervention began for the next group. Midway through the study, groups were rearranged if necessary based on progress monitoring data. Gyovai et al. concluded that intervention effects may have been greater if instruction began sooner in the school year and extended for a longer duration. Because of staggered start times and regrouping, students received between 12 and 50 instructional sessions. Results indicated that all students made progress in their phonemic and word reading skills during the intervention phase but with considerable variation in the degree of progress. Gyovai et al. noted that continuous monitoring of progress allowed for important instructional adjustments during the intervention phase.

Musti-Rao and Cartledge (2007) selected eight students for participation in their study. Students were either in the at-risk or some-risk category according to reading screening data on DIBELS. In addition to progress monitoring measures, Musti-Rao and Cartledge administered pretest and posttest, standardized measures of word reading for the eight students in their study and created their own curriculum-based assessment to more closely measure student progress on specific skills taught throughout the intervention phase. During the intervention phase, one student was moved to his own group due to significant progress beyond the other group members. Similar to Gyovai et al. (2009), all students achieved an accelerated rate of progress on phonemic and word reading skills during the intervention phase as compared to baseline. Musti-Rao and Cartledge reported that two students demonstrated a delayed response to intervention, in which gains were not seen until four weeks of instruction. At the end of the study, two students remained in the at-risk category and one remained in the some risk category. The authors concluded that at-risk students in kindergarten can benefit from explicit and systematic instruction in early reading skills. However, a couple of the students in this study appeared to need more intensive supports.

Looked at collectively, the studies reviewed in this section inform the literature in a number of respects. First, early intervention in reading clearly leads to successful outcomes for many students identified in kindergarten as at risk for reading failure. Second, a subgroup of these students does not respond sufficiently to these intervention supports. Third, a systematic approach to increasing instructional intensity is lacking. Excluding the multiple baseline studies, only the study by O'Connor (2000) attempted to make instructional adjustments during the school year based on student responsiveness. Fuchs et al. (2002) and Scanlon et al. (2005) did

not evaluate instructional responsiveness until the beginning of the following school year. Next, I provide a review of kindergarten intervention studies that were more intensive in nature.

More Intensive Kindergarten Interventions

As compared to the studies discussed in the previous section, the studies discussed here consisted of interventions delivered in a more intensive format according to such factors as group size, frequency of instruction, duration of sessions, and number of weeks of instruction (Coyne et al., 2013; Hagan-Burke et al., 2011; Torgesen et al., 1999). Two of these studies are from recent publications, perhaps suggesting they were conducted to address the need for more intensive kindergarten intervention studies. All studies reviewed were experimental in nature and utilized pretest-posttest comparison group designs. Whereas the Torgesen et al. (1999) study was conducted in kindergarten through second grade, Hagan-Burke et al. (2011) and Coyne et al. (2013) implemented their interventions in kindergarten only. Additionally, the Coyne et al. study is a replication of the Hagan-Burke et al. study.

Students in the Torgesen et al. (1999) study were selected through a two-part process. In January of kindergarten, children were initially screened on an uppercase letter naming task. Students who scored in the bottom 30% were given a phonemic elision task, a serial naming of numbers task, and a measure of vocabulary. Children with an estimated Verbal Intelligence score above 75 who obtained the lowest combined scores on the letter naming task and the phoneme elision task were included in the study (n=180). Interesting here is that Torgesen et al. used a combined score to identify students at risk, but included both alphabet knowledge and phonological awareness measures. In Hagan-Burke et al. (2011) and Coyne et al. (2013), teachers and principals were asked to review existing school data to identify potential candidates for the study. Five to eight students from each classroom were then nominated for supplemental

reading support. Nominees were then screened on a letter naming fluency and a sound matching test. By nominating a limited number of students from each classroom prior to initial screening, it is possible that some students with greater needs did not participate in the screening. Instead of combining scores as in Torgesen et al., students who scored at or below the 33rd percentile on letter naming fluency or below the 37th percentile on a sound matching task were selected to receive intervention.

In contrast to the studies reviewed in the previous section, interventions were delivered with more instructional intensity. In Torgesen et al. (1999), students received one-to-one instruction four times a week for 20 minutes from January to the end of kindergarten. Interventions in Hagan-Burke et al. (2011) and Coyne et al. (2013) were delivered in groups of three to five students for 30 minutes per day for 21 weeks. Intervention specifics and results for all three studies are described next.

Torgesen et al. (1999) randomly assigned to one of four conditions: an embedded phonics condition (EP), a phonological awareness plus synthetic phonics condition (PASP), a regular classroom support (RCS) condition, or a no-treatment control (NTC). Children in the PASP condition received instruction in phonemic awareness, letter-sound correspondences, decoding, spelling, and reading and writing of connected text from the Lindamood Bell LIPS program. Children in the EP condition initially received instruction in learning to recognize small groups of whole words, instruction in letter-sound correspondences in the context of sight-word instruction, writing the words in sentences, and reading the sentences that were written. Children in the RCS condition received tutoring aligned to their regular classroom reading programs which varied across schools. Through these different experimental conditions, Torgesen et al. were interested in investigating the effects of differences in the balance between word and text

level instruction for children with phonological weaknesses (as measured by the letter naming and phoneme elision tasks), on a variety of reading outcome measures. Torgesen et al. noted that the children selected for the study were among the bottom 12% in phonological processing skill.

Based on the results of lesson observations, it was estimated that students in the PASP and EP conditions received instruction in phonological awareness, letter-sound correspondences, and phonemic decoding 74% and 26% of the time respectively, direct instruction and practice in sight-word instruction 6% and 17% of the time respectively, and reading and writing connected text 20% and 57% of the time respectively. Tutors in the RCS condition classified their instructional time as 24% sight word activities, 24% phonics activities, 9% spelling activities, and 43% meaning-emphasis activities. In general, students in the PASP condition spent 80% of their time on word-level instruction and 20% on text level activities, and students in the EP condition spent 43% and 57% of their time on word and text level activities respectively.

Outcome measures of word reading skills and phonological awareness were given at the end of each grade, as well as in the middle of first and second grade. A more extensive battery was also administered at the end of second grade that included measures of reading comprehension, verbal ability, and behavior.

Students in the PASP group consistently obtained the highest scores on reading outcome measures and were close to the 50^{th} percentile on word reading skills and on the low end of the average range on reading comprehension measures. Students in the PASP condition showed a statistically significant difference over students in the EP condition in phonological awareness, phonemic decoding, and word reading. In the analyses involving all four conditions, a statistically reliable effect was found for students in the PASP condition on word-level reading

skills but not comprehension. However, 24% and 21% of children in PASP condition scored one standard deviation (SD) below average in phonemic reading skills and real word reading ability respectively. No statistical differences were found in word-level skills between students in the RCS and EP groups.

A statistically significant difference in retention rates was also found across conditions for kindergarten and first grade. Retention rates for the NTC, RCS, EP, and PASP conditions were 41%, 30%, 25%, and 9% respectively. A statistically significant difference was also found in the referral rate for special services when just the PASP and EP conditions were compared. Whereas 42% of the students in the EP condition were referred for special services, only 18% in the PASP were referred.

Other assessment results revealed that rapid naming, home background, and classroom behavior ratings were found to be the biggest predictors of word reading skills. Rapid naming and classroom behavior ratings were found to be the biggest predictors of reading comprehension. The phonological variables contributed more to the prediction of growth in word attack skills than word identification skills. General verbal ability was not a significant predictor of word reading skills but was a significant predictor of reading comprehension.

Torgesen et al. (1999) concluded that interventions for students with phonological processing weaknesses must include explicit and intensive instruction in phonemic awareness and decoding skills, as well as enough time to build comprehension skills. They noted, “To be maximally effective, early-intervention programs need to contain a carefully orchestrated mix of instruction to help children construct the meaning of text as well as to read words accurately and fluently” (p. 580).

The Torgesen et al. (1999) study is important to the literature base in terms of exploring the mixture of code- and meaning-focused intervention components. In this study, a significant dose of code-focused instruction appeared to be important to later, successful literacy outcomes, at least for students with the lowest letter naming and phonemic awareness abilities. Specifically, the students in the PASP condition received approximately 16 minutes of word-level instruction per 20-minute session as compared to approximately 8.6 minutes in the EP condition. Although students in the PASP condition consistently demonstrated the highest response, their low average scores on reading comprehension measures suggests additional instructional time devoted to meaning-focused activities. Since students received intervention across three grade levels, what becomes less clear is when this shift should occur. Perhaps the 80% and 20% mixture of code- and meaning-focused instruction respectively in a 20-minute timeframe is appropriate for kindergarten intervention but should gradually shift over time to include a large meaning-focused component as students progress from prereaders to a beginning stage of reading. Other possibilities include beginning intervention services with a slightly less discrepant gap in code- and meaning-focused instruction or increasing the total instructional time for those students with the highest risk level to 30 minutes per day. A feasibility issue also arises in this study with one-to-one groupings for all students across three academic years.

Hagan-Burke et al. (2011) examined if explicit, code-based instruction (ESC condition) or typical school intervention (TIP condition) moderated the strength of the above factors. Participants attended schools in south-central Texas and Connecticut. Eleven of 12 schools received Title I funding with 50% to 81% of students enrolled in free and reduced lunch. The ESC condition consisted of instruction from the Early Reading Intervention (ERI) program. ERI, as described in Simmons et al. (2011), consists of 126 lessons, delivered in 30 minutes,

with seven activities of three to five minutes each. In the first 15 minutes of the lesson, students receive instruction in phonological awareness and alphabetic understanding. In the second 15 minutes, skills from the first 15 minutes are integrated into writing and spelling activities. The program is further divided into four parts. The focal points of each part change as students progress through the program. Part I consists of learning the letters names and sounds of 11 letters and the phonemic skills of first and last sound isolation. Part II consists of learning five new letter names and sounds and phonemic blending and segmenting using letter tiles. Part III consists of learning six more letter names and sounds focusing on the decoding of vowel-consonant and consonant-vowel-consonant words. Oral segmenting and blending is integrated with the decoding of regular words and irregular sight word reading is introduced. Part IV focuses on integrating alphabetic skills and irregular word reading to read sentences and short storybooks.

Students in the TIP condition received whatever intervention-based services their school offered. For both groups, interventions were delivered in groups of three to five students for 30 minutes per day for 21 weeks. Forty percent of students in the ESC condition and 38% of students in the TIP condition received intervention in a pullout setting. Coyne et al. (2013) replicated the study described by Hagan-Burke et al. (2011). The researchers wanted to see if results from the ERI program would replicate to a Florida School District with a more consistent, coordinated, and systematic approach to early literacy instruction. As in the original study, the ERI intervention was compared to school-designed interventions (SDI), and both interventions were again implemented daily for 30 minutes, five days per week, and in groups of three to five students.

For the original study, results for participants in the ESC and TIP conditions are described by Simmons et al. (2011) and are as follows. Both interventions were effective for the majority of students. More specifically, Simmons et al. reported that students in the ESC condition obtained higher mean scores on all measures except comprehension. Statistically significant differences were obtained for measures of alphabet knowledge, word attack, and phonemic awareness with effect sizes ranging from .26 to .51. Interestingly, the magnitude of effect sizes generally corresponded with the scope and sequence of the ERI program. In other words, the higher effect sizes corresponded to the skill areas that received the most instructional focus.

Additionally, fewer percentages of students in the ESC condition remained below the 15th percentile on all measures except the comprehension subtest. Across a range of phonemic and word reading measures, students in the TIP condition were two to four times more likely to perform below the 15th percentile. Students who scored at or above the 30th percentile in outcome measures were considered to be no longer at risk. On the WRMT-R/NU Word Attack subtest, 1.8% of students in the ESC condition and 3.3% of students in the TIP condition were nonresponders. According to the DIBELS NWF subtest, 36.6% and 44.7% of students in the ESC and TIP conditions respectively were nonresponders. In general, normative-referenced measures such the Woodcock Reading Mastery Tests-Revised/Normative Update (WRMT-R/NU; 1998) Word Attack subtest identified fewer students as nonresponders than timed measures such as the NWF (Simmons et al., 2011).

Unlike in the original study, Coyne et al. (2013) found no statistical differences between the ERI and SDI conditions on any outcome measures, and a trend actually existed favoring the SDI condition. However, the ERI condition in the initial and replication studies had similar

posttest results with a statistically significant difference found only for phonemic segmentation fluency in favor of students in the Florida group. When comparing the two SDI conditions, statistically significant differences were found favoring the SDI replication group on "measures of phonemic awareness, letter-sound knowledge, nonsense word fluency, and word identification, with effect sizes ranging from .24 to 1.06" (Coyne et al., 2013, p. 19).

In the initial study conducted in Connecticut and Texas, participating school districts were characterized by Coyne et al. (2013) as providing a less coordinated and more individual approach to kindergarten reading instruction and intervention. The Florida school district, on the other hand, was characterized as having a more coordinated, systematic, and consistent approach. For example, all students in the Florida district received 90 minutes of core instruction per day from a comprehensive core reading program (Harcourt Trophies; Beck, Farr, & Strickland, 2007). In addition, all teachers who implemented the school-designed interventions had received extensive professional development in evidence-based interventions and had experience delivering kindergarten intervention supports. Coyne et al. described these supports as including the strategic integration of phoneme awareness and alphabetic instruction. Based on observation data, Coyne et al. noted that teachers in the replication SDI condition emphasized phonological blending and segmenting, sight word work, reading connected text, and writing sounds and words more so than SDI teachers in the initial study. Teachers in the initial study had a larger emphasis on phonological skills at the word and syllable level. The authors concluded that instructional context may have been an important factor in explaining the differences in outcomes across these two studies.

An interesting aspect of both the initial and replication study is the documentation of instructional focal points, quality of instruction, and dosage amounts across conditions. By

systematically documenting the instructional focal points, the reader is more informed as to particular aspects of intervention content that potentially explain differences in outcomes. Although one cannot conclude that differences in outcomes across the two SDI conditions were a result of the instructional differences mentioned above, they become more meaningful when the quality of instruction and dosage data is also considered. In the two SDI conditions, quality of instruction data is virtually identical. Additionally, students in the initial SDI condition received on average 106 days of instruction as compared to approximately 88 days of instruction for students in the replication study. In light of the speculation raised by Simmons et al. (2011) pertaining to the limited sight word instruction received by ERI students in the initial study, it is interesting that students in the SDI replication condition, who likely received the highest percentage of sight word instruction, achieved the highest mean score in word identification skills.

Hagan-Burke et al. (2011) also investigated the effect of student, teacher, and setting factors on the participants in their study. In terms of student variables, results indicated that entry-level alphabet knowledge had a statistically significant positive effect on decoding (y = .59) and phonemic awareness (y = .41). Students in both intervention groups with higher alphabet knowledge tended to score higher on both decoding and phonemic outcome measures. Rapid automatic naming, (RAN), as measured by a rapid object naming test, influenced decoding irrespective of instructional condition (y = .28). In other words, students with higher RAN scores tended to score higher on end-of-year decoding measures. However, RAN performance was a statistically stronger predictor of phonemic awareness in the TIP condition only, where students with lower preintervention RAN scores tended to score lower on the composite phonemic measure. Hagan-Burke et al. (2011) suggested that perhaps the explicit phonemic

instruction within the ERI condition was sufficient to moderate RAN's influence on phonemic outcomes. There was no significant influence of entry-level vocabulary or sound matching on decoding or phonemic awareness outcomes. As vocabulary scores were less predictive of phonemic awareness and decoding outcomes, the authors proposed that vocabulary knowledge "may be a better predictor of outcomes more closely related to vocabulary knowledge and performance during later grades" (p. 273). The authors further suggested that composite scores from the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999) may have provided a more complete picture of students' phonemic awareness skills as opposed to using the Sound Matching subtest only.

In terms of teacher variables, only quality of instruction had a statistically significant effect on decoding (y = .36) and was a stronger predictor for children in the TIP condition. In general, a one standard deviation increase in a TIP teacher's instructional quality score was connected with .36 standard deviation (SD) increase on decoding scores. Hagan-Burke et al. (2011) hypothesized that quality of instruction for students in the TIP conditions had a more significant effect on decoding because of the less structured nature of interventions in the TIP condition as compared to the ESC condition. On average, teachers in the ESC condition tended to score about 10% points higher on teacher quality. Children in both conditions tended to score the same on decoding and phonemic measures "regardless of their teachers' years of experience and level of phonemic knowledge" (p. 272).

In regards to setting variables, intervention delivery (i.e. pullout versus in-class intervention), did not influence decoding and phonemic awareness outcomes, but group size had a statistically significant influence on phonemic awareness. For the TIP condition, students in larger groups tended to score lower on end-of-kindergarten phonemic awareness measures. For

every one fewer student per group, students scored about .39 SD units higher on phonemic awareness outcomes. This suggests that smaller group sizes may be important for kindergarten students with low phonemic awareness skills. As students started off in groups of three to five students, perhaps the most at-risk students would be better served in groups of two to three students. It may be that students in smaller groups receive more opportunities to respond and additional corrective feedback important to their growth in phonemic awareness. Interestingly, group size did not influence outcomes in the more structured ESC condition, where teachers tended to score higher in quality of instruction ratings.

Overall, Hagan-Burke et al. (2011) concluded that "models containing quality of teacher's instructional practices and setting variables explained at least as much variability in kindergarten reading outcomes as the model with students' entry-level language and literacy skills" (p. 275). As with the Fuchs et al. study (as described in Al Otaiba & Fuchs, 2006), instructional quality influenced student outcomes.

In their discussion of intervention findings, Simmons et al. (2011) make a number of insightful comments. First, the authors noted that sight word instruction does not occur until the latter half of the ERI program and could explain why statistically significant effects were not found in word identification between students in the two different intervention conditions. If sight word instruction was introduced earlier, students at risk for reading failure may receive the additional practice needed for more automatic word identification. Additional research is needed to explore this premise. In implementing additional sight word instruction, it would be important not to compromise the amount of phonemic decoding instruction as the highest effect size was found on the Word Attack subtest (i.e. a measure of phonemic decoding).

Second, Simmons et al. (2011) addressed the findings related to passage comprehension outcomes. As previously stated, no significant difference was found between groups on the passage comprehension measure. The ERI intervention in this study had little emphasis on reading comprehension. Even though more attention was typically given to reading comprehension in the TIP condition, the mean score for the TIP group was lower than that of students in the ERI condition. Simmons et al. concluded that additional research is needed to "explore the benefits of explicitly targeting reading comprehension in kindergarten interventions" (p. 223).

Third, Simmons et al. (2011) suggested that future research include a more responsive schedule of instruction based on formative assessment. With this type of intervention format, the pace of instruction can be altered in response to individual student progress and therefore, meet the needs of students who would benefit from either a faster or slower pace. With groups of three to five students, however, regrouping of students based on their responsiveness would likely be necessary to create homogeneity across groups. Lastly, Simmons et al. observed that the intensity of intervention delivery in this study may have been an important component in the accelerated learning of students in both conditions. This study clearly differs from other kindergarten intervention studies reviewed thus far in that it included all three of the following elements of instructional intensity: daily instruction, sessions of a longer duration (i.e. 30 minutes), and intervention delivered for at least 20 weeks. Also important to note is that instruction began earlier in the school year, as opposed to in January. In the next section, studies that compare more and less intensive kindergarten interventions are reviewed.

Comparisons of More and Less Intensive Kindergarten Interventions

A number of studies have been conducted to compare the effects of more and less intensive kindergarten interventions (Al Otaiba, Schatschneider, & Silverman, 2005; Lennon & Slesinski, 1999; Simmons et al., 2007). These studies are described and analyzed, as well as synthesized with findings of other studies reviewed.

Lennon and Slesinski (1999) conducted a kindergarten intervention consisting of both code- and meaning-focused instructional components. Three hundred thirty students were screened on a letter naming test consisting of all uppercase and lowercase letters arranged in random order. Using these results, the following three groups of students were created: a low-scoring group, a mid-range group, and a high-range group. The low-scoring group, consisting of the 16 lowest-scoring students from each of the five participating schools (n = 80), was randomly assigned to one of two groups. One group received intervention during the first ten weeks and the second group during the next ten weeks. Twenty-four students from the first ten-week group continued with the intervention during the second ten-week period. Fifty-six mid-scoring students were randomly selected to receive the intervention, 40 students during the first ten weeks and 16 during the second ten weeks. Finally, a comparison group was created from a random selection of 40 students from the high-scoring group, defined as students scoring in the top 15% on the letter naming task. Students in the high-scoring group were administered the same dependent measures as students receiving the intervention. Dependent measures consisted of tests of alphabet knowledge, phonemic awareness, decoding, and concepts of print.

Tutoring was delivered in groups with a teacher-to-student ratio of one-to-two for 30 minutes five times a week. Lennon and Slesinski (1999) noted that the one-to-two groupings

were a compromise among stakeholders between one-to-one instruction and smaller group instruction of three to five students. They believed this lower ratio would allow for more individualized attention that students with severe reading problems may need to become successful readers. Lennon and Slesinski commented further that the one-to-two arrangements would also extend previous literature of intensive intervention using a one-to-one arrangement.

Students were paired on their initial scores on the letter naming task, and groups were rearranged if students progressed at differential rates. Students received 15 minutes of instruction in phonemic awareness, letter names, letter sounds, the alphabetic principle, and sight word reading. Fifteen minutes was used for shared-reading activities. Thus, this study had a 50/50 balance of code- and meaning-focused instruction. Lennon and Slesinski (1999) commented that instruction in both of these domains is necessary for students at risk for reading failure to build competent reading skills and that perhaps the best approach would be in the context of integrated activities. Tutors were also encouraged to integrate concepts and skills from students' regular classroom instruction into their sessions.

Results showed that low-range students in the first ten-week group outperformed low- and mid-range students in the waiting group, with a statistically significant difference found on a number of reading outcome measures. In addition, students in the low group who received 20 weeks of instruction surpassed students in the low group who received 10 weeks of instruction and functioned similarly to the mid-range students who received 10 weeks of instruction and the comparison group of high-range students. In other words, students in the low-scoring range who received 20 weeks of instruction improved their relative standing on outcome measures, whereas those who received ten weeks of instruction did not. Lennon and Slesinski (1999) suggested that these results support the idea that initial screening can be used to determine dosage level.

Several features of this study are important for further discussion. First, a number of other studies reviewed placed students either into groups of three to five students (Coyne et al., 2013; Hagan-Burke et al., 2011; O'Connor, 2000; Scanlon et al., 2005) or in a one-to-one format (O'Connor, 2000; Torgesen et al., 1999). Lennon and Slesinski (1999) contended that a consideration of student need may allow for more intensive ratios of instruction for longer time periods. As school-based resources are limited, such differentiation of instructional intensity based on level of need may be a practical solution for many school districts. For students just entering kindergarten, those with the highest needs could theoretically be placed in more intensive interventions of longer durations and smaller groups, perhaps two to three students per group. Those with more moderate needs could receive a less intensive intervention perhaps for a shorter duration and larger groups of three to five students. Evidence from other studies reviewed supports such an arrangement (Fuchs et al., 2002; Hagan-Burke et al., 2011; O'Connor, 2000; Scanlon et al., 2005; Simmons et al, 2011; Vellutino, et al., 2006).

Second, tutors were asked to keep daily records of student progress. This information was used to create more homogeneous groups that allowed tutors to adjust their instruction to the targeted needs of their students. Interestingly, tutors used student responsiveness to their daily instruction to make these determinations as opposed to a more formalized measure. No other studies with experimental, pretest-posttest group designs reviewed thus far have included a component allowing for the rearrangement of groups based on individual student need. As noted in Vellutino et al. (2006), distinguishing between students who are less- and more-difficult to remediate can be difficult based on pretest measures. Even when considering only those students with the most intensive needs, it is likely that response to intervention will differ among students. Keeping students in the same group with disparate responses will obviously complicate pacing of

instruction. Recall that this issue was discussed in Simmons et al. (2011), and the authors of that study suggested the need for future research that includes a more responsive schedule of instruction based on formative assessment. This was clearly a component in the instructional design of Lennon and Slesinski (1999).

A third point of discussion pertains to results on outcome measures. As previously discussed, Lennon and Slesinski (1999) reported that students in the low-ranging group who received 20 weeks of instruction performed similarly to the mid-range group and the high-range comparison group on a range of measures. However, there appears to be a large discrepancy in mean scores between both the 20-week low-range and 10-week mid-range groups and the high-range comparison group on the sight word measure. The authors did not discuss this result specifically and the statistical significance of this difference is unreported. However, this result appears to parallel that of other studies in which word identification scores are among the most difficult to boost for early readers who struggle. This again suggests that this component of code-focused instruction receive more attention.

Al Otaiba et al. (2005) conducted a year-long study investigating the effectiveness of the Tutor Assisted Intensive Learning Strategies (TAILS) program. The intervention was delivered by community tutors to students identified at risk for reading difficulties. Two hundred forty-three students in four, low-performing, urban schools were screened in October with the DIBELS (2002) Letter Naming Fluency (LNF) and Initial Sound Fluency (ISF) subtests. At-risk status was defined as correctly identifying less than two letter names or three initial sounds on the LNF or ISF tests respectively, which corresponded to the lowest 30th percentile.

One hundred five students met the criteria, but intervention could only be delivered to 76 students because of limited resources. Consequently, the seventy-six students with the lowest pretest scores were matched into triads based on LNF scores and randomly assigned into one of three treatments: TAILS four days a week (TAILS-4), TAILS two days a week (TAILS-2), and a comparison group. TAILS lessons were about 30 minutes and occurred outside of the 90-min literacy block.

TAILS includes all five components identified by the National Reading Panel (2000): phonemic awareness, phonics, fluency, vocabulary, and comprehension. More specifically, TAILS lessons consisted of the following: five minutes of phonological awareness activities; five minutes of word-building activities; ten minutes of decoding activities; and ten to fifteen minutes of shared book reading activities. Phonological awareness instruction involved initial sound, rhyming, blending, and segmenting activities. Word-building instruction involved the integration of phonological awareness and letter symbols using both manipulatives and letter tiles to represent the smaller parts of words. Letter tiles were used only for those letters that had previously been taught. Decoding instruction began with first learning to read some letter sounds, then learning to blend sounds into words as new letter sounds were progressively introduced. Other decoding activities involved sight word instruction and eventually the reading of simple sentences. Shared book reading activities utilized dialogic reading practices (Bus, van IJzendoorn, & Pellegrini, 1995; Lonigan & Whitehurst, 1998), incorporated vocabulary instruction based on Beck, McKeown, and Kucan (2002), and included a range of books from a number of genres. Student also read simple decodable books during this time. The comparison group was read aloud to with the same stories from TAILS two times per week for 20 minutes.

Students were administered a number of pretests measuring phonemic awareness, word reading, reading comprehension, and vocabulary. Statistically significant differences were found in the TAILS-4 group as compared to both the TAILS-2 and the comparison groups for measures of word identification, passage comprehension, and a basic reading skills cluster, with effect sizes of .79, .90. and .83 respectively. A statistically significant difference was also found for the TAILS-2 group over the comparison group on a measure of phonemic blending, with an effect size of .68.

According to Al Otaiba et al. (2005), these results support four days a week of the TAILS intervention program for at-risk readers over simply two days a week. Additionally, two days a week appeared to be insufficient to produce significant gains beyond those made by students in the comparison condition. Al Otaiba et al. hypothesize that the success of the TAILS intervention in this study was in part due to tutor training, the high degree of fidelity of implementation (> 97% during the study), the clear scope and sequence, and a design that included critical aspects of early literacy instruction. Even so, Al Otaiba et al. reported that some students were nonresponsive to the tutoring. Additionally, no statistically significant difference was found in receptive vocabulary growth. The authors suggested future use of a more sensitive measure of expressive language and more explicit vocabulary instruction. Al Otaiba et al. further identified the lack of individualized programming as a limitation to the study.

Although students who received instruction four days a week demonstrated significant growth on a number of measures, the percentage of students who were nonresponsive is not reported. This information would be useful in further evaluating the effectiveness of the program. The number of students per group is also not specified. Results generally support the amount of time devoted to different code-focused elements (i.e. 20 minutes per session). In

addition to a .79 effect size for students in the TAILS-4 over students in the other two conditions on word identification skills, the mean pre- and posttest standardized scores for TAILS-4 students improved from 85.29 to 100.50. Additionally, the mean posttest on a measure of word attack skill for the TAILS-4 group was 101.17. From the authors' description, phonemic decoding and sight word instruction appear early in the program. Perhaps this early emphasis helps to explain mean scores right around the 50th percentile for both word identification and word attack skills.

Results for the meaning-focused components are less clear. Although an effect size of .90 was found for students in the TAILS-4 group on comprehension skills, the mean pre- and posttest scores were virtually identical (i.e. 94.50 and 94.75). In other words, these students maintained their position relative to their peers on this particular measure. On the other hand, mean scores for students in the other two conditions dropped from pre- to posttest, indicating the instruction they received was not able to maintain these students at pre-treatment levels.

Simmons et al. (2007) examined differences in design specificity and instructional time of three kindergarten interventions. One intervention consisted of 30 minutes of highly specified, code-based instruction (30/H). The second intervention consisted of 15 minutes of code emphasis that matched the first 15 minutes of the first intervention and 15 minutes of vocabulary and listening comprehension instruction through a storybook activity (15/H + 15). The third intervention consisted of 30 minutes of code-emphasis instruction but of moderate specificity as compared to the first (30/M). Interventions were supplemental to the 45-60 minutes of daily reading instruction provided in their regular classrooms. Instruction was delivered in groups of five or fewer students from November to mid-May and students averaged 108 days and 54 hours of instruction.

One hundred sixteen kindergarten students from seven schools that all received Title I funding were screened with DIBELS (Kaminski & Good, 1996) LNF and Onset Recognition Fluency (OnRF). Students were included in the study if they scored at or below the 25th percentile of district scores. Dependent measures included tests of phonemic awareness, word reading, and spelling.

Students in the 30/H condition received systematic instruction in phonemic awareness, alphabet knowledge, letter writing, and spelling. Activities in the first 15 minutes included first and last sound isolation, phonemic blending and segmenting, letter-name and letter-sound identification, letter-sound blending, irregular word reading, and sentence reading. The second 15 minutes included the writing of letter sounds, integrated phonological and alphabetic tasks, and spelling. Letter writing did not include specific instruction of letter formation. Essential components to the highly-specified condition included teacher scaffolding, strategic integration of skills, systematic review, explicit instructional language, specific procedures for corrective feedback, and extended feedback for difficult skills.

Students in the 15/H + 15 group received the same code-focused instruction of the first 15 minutes of the 30/H condition. The second 15 minutes included instruction of vocabulary, story structure, and story retell embedded in shared book reading that included repeated reading of stories, targeted vocabulary instruction with multiple exposures to words and systematic review, and dialogic discussion.

Students in the 30/M group also received 30 minutes of code-focused instruction based on the Sounds and Letters component of Open Court Reading 2000 (Adams et al., 2000). Simmons et al. (2007) characterized the activities from Sounds and Letters as moderately

specified because they were broader in scope (e.g. included a broader range of phonological awareness activities), included no time allocations per task, varied in the explicitness of teacher language and direction, and included time for singing songs, reading poetry, and playing language games to reinforce new skills.

In terms of instructional time, results indicated 30 minutes of highly-specified instruction was comparable to 15 minutes of highly-specified instruction for increasing phonemic awareness skills in initial sound isolation and phonemic segmentation but significantly more effective in increasing levels of automatic retrieval and production of handwritten letters. Thirty minutes of highly-specified instruction was also significantly more effective in increasing levels of phonemic decoding, word attack skills, word identification skills, and spelling proficiency of only those kindergarteners who were most at risk at the start of the intervention on the letter naming fluency and developmental spelling assessments. Most at risk corresponded to fewer than four letter names per minute or fewer than 13 correctly spelled letters in a list of words. This finding lends support for more intensive code-focused instruction for those students with the lowest entry-level alphabetic skills. For students in this study with higher entry-level alphabetic skills, 15 minutes of highly specified code-focused instruction was sufficient to adequately boost their word attack, word identification, and spelling skills.

In terms of design specificity, 30 minutes of highly-specified instruction was comparable to 30 minutes of moderately-specified instruction for increasing phonemic awareness skills in initial sound isolation and phonemic segmentation, significantly more effective in increasing levels of phonemic decoding, spelling fluency, automatic retrieval and production of handwritten letters, and significantly more effective in increasing levels of word attack and word identification skills for students most at risk on entry-level letter naming fluency.

The results of this study further support the effectiveness of explicit, code-based instruction delivered in an intensive format (i.e. daily instruction for 30 minutes in duration). This type of instruction was particularly important for those students with the lowest scores in letter naming fluency. In their conclusion, Simmons et al. (2007) raise two important questions: 1) What portion of time in a 30-minute code-focused intervention should be allocated to phonemic, alphabetic, and letter writing and spelling? and 2) Can the 30-minute timeframe be compressed? Additional research is needed to answer these questions.

Collectively, the articles in this section inform kindergarten reading intervention in a number of respects. The results of all three studies (Al Otaiba et al., 2005; Lennon & Slesinski, 1999; Simmons et al., 2007) support more intensive interventions for kindergarten students identified as at risk for future reading difficulties. Interestingly, all three studies compared an increase in instructional intensity in different ways. In respect to the Lennon & Slesinski (1999), researchers compared the total length of instruction in terms of number of weeks. Al Otaiba et al. (2005) compared intensity in terms of frequency of instruction (i.e. the amount of days per week). Simmons et al. (2007) investigated the amount of time per session as well as the degree of instructional specificity.

In general, findings revealed that 20 weeks of instruction was superior to 10 weeks, four days a week was superior to two days a week, 30 minutes of code-focused instruction was superior to 15 minutes, and highly-specified instruction was superior to moderate specificity. Interventions in both the Lennon and Slesinski (1999) and Simmons et al. (2007) studies were scheduled for at least four days a week. Al Otaiba et al. (2005) reported that four days a week appears to be necessary. Code-focused instruction occurred between 15 and 30 minutes across studies and groups, and meaning-focused instruction occurred between zero and 15 minutes

across studies and groups. In relation to instructional design specificity, Lennon and Slesinski reported the use of direct instruction techniques for decoding and writing instruction. Al Otaiba et al. characterized the decoding portion of TAILS lessons as being based on principles of direct instruction. Activities in the phonological awareness and decoding portions of the lessons also adhered to the model-lead-test format in which tutors first modeled the skill and then led students through repeated practice.

All of the studies reviewed thus far included multiple code-focused components, and some included meaning-focused components as well. These studies have informed the literature in respect to instructional intensity. They have also informed the literature in terms of important components of kindergarten reading intervention for students at risk, such as instruction in phonological awareness, alphabetic knowledge, decoding, vocabulary, and comprehension. Taking into consideration their instructional focus, the results of several studies suggest that the phonological skills of first and last sound isolation, phonemic blending, and phonemic segmentation are essential components to include (Al Otaiba et al., 2005; Coyne et al., 2013; Scanlon et al., 2005; Simmons et al., 2007; Simmons et al., 2011; Torgesen et al., 1999). They do not, however, give a complete picture of the organization and sequencing of alphabet knowledge instruction. For example, although virtually all of the studies included letter name and letter sound instruction, an ideal scope and sequence of alphabet instruction and whether or not to include both uppercase and lowercase letters remains unclear. When considering students with the most intensive needs, some of whom do not respond to instruction, this information is particularly relevant. Consequently, studies that inform the knowledge base of alphabet knowledge instruction for kindergarten students are discussed next.

Components, Organization, and Sequencing of Alphabet Knowledge Instruction

A number of studies have been conducted that inform researchers and practitioners in regards to alphabet knowledge instruction (Evans, Bell, Shaw, Moretti, & Page, 2006; Jones, Clark, and Reutzel, 2013; McBride-Chang, 1999; Piasta & Wagner, 2010; Turnbull, Bowles, Skibbe, Justice, and Wiggins, 2010). Some studies have found that letter names can facilitate letter-sound knowledge (Evans et al., 2006; McBride-Chang, 1999; Treiman, Tincoff, Rodriguez, Mouzaki, & Francis, 1998). McBride-Chang (1999) investigated the letter-name and letter-sound knowledge of 91 kindergarten students who entered school as prereaders. Students were measured at four points in time over a 15-month period. Results indicated that letter-name knowledge predicts both later letter-name and letter-sound knowledge, whereas letter-sound knowledge predicts later letter-sound but not letter-name knowledge. Children tended to learn more letter names initially than letter sounds. McBride-Chang also found that letter names are supportive of learning letter sounds. The easiest letter sounds to learn were those with the letter sound as the initial phoneme in the letter name, such as b, d, t, and k, a finding consistent with Evans et al. (2006). Evans et al. further found that kindergarteners in their study knew letters whose name ends with the sound of the letter (e.g. f, l, n) better than letters that do not have their sound in their name (e.g. h, w, y). Children knew on average 89.1% of uppercase letters, 74.9% of lowercase letters, and 57.9% of letter sounds. However, children knew 69.9% of letter sounds when they also knew the name of the letter. Generally, children knew both the uppercase and lowercase letter name for those letters with a very similar upper- and lowercase form. No relationship was found between letter name knowledge and the order in which the letter appears in the alphabet.

Turnbull et al. (2010) found that preschoolers in their study were 16 times more likely to know a lowercase letter if they knew the corresponding uppercase letter. They further found that uppercase and lowercase similarity, letters in a student's name, and letter frequency in printed English predicted children's lowercase letter knowledge. Jones et al. (2013) discussed a number of factors and characteristics that can be considered when creating an alphabet knowledge scope and sequence. One such factor is distinct visual feature advantage. This advantage suggests separating letters with similar visual features in the instructional sequence, such as C and G. Additionally, the letter C, with fewer features than the letter G, would be introduced first. When G is introduced, its visual features can be compared to the letter C, which students are already familiar with. Another factor discussed by Jones et al. is developmental phoneme acquisition order, which suggests introducing letters with developmentally easier to articulate sounds earlier in the instructional sequence.

Piasta and Wagner (2010) conducted a recent meta-analysis of alphabet knowledge instruction. Overall, studies had moderate effects on alphabet knowledge as measured by untimed assessments involving letter name recognition or production, untimed letter sound recognition or production, letter naming fluency, and letter sound fluency. However, there was no effect on letter naming fluency outcomes when these were parsed out from other outcome measures. When measured, small effects were found on reading skills immediately following intervention. Follow-up assessments given two to twelve months later typically demonstrated no effects.

Studies included in Piasta and Wagner's (2010) meta-analysis that consisted of multiple early literacy components (e.g. letter name or letter sound instruction combined with phonological awareness instruction) had higher effect sizes, perhaps lending support to the

reciprocal relationship among early literacy skills. Interestingly, studies that included letter name instruction only produced positive impacts on letter sound learning. Letter name knowledge was impacted by both size of group and amount of instruction; smaller group sizes and longer amounts of instructional time produced larger impacts.

Piasta and Wagner (2010) concluded that alphabet learning may require significant amounts of time and repeated practice. They surmised that overall moderate effect sizes "may be an authentic representation of our current ability to foster alphabet knowledge development during early literacy instruction" (22), and in order to increase the alphabet knowledge of at-risk students to desirable levels, more intensive, explicit alphabet knowledge instruction may be necessary. Piasta and Wagner further hypothesized that "letter name or sound instruction does not readily transfer to reading and spelling without instruction and practice in using alphabet knowledge for these purposes" (24). More research is needed to explore potential moderators of intervention effect, such as letter sequence, instructional materials, and instructor training.

Together, these studies can help guide alphabet knowledge instruction for students who have the least alphabet knowledge and who may have difficulties in alphabet knowledge acquisition. These findings suggest that beginning with uppercase letters may be beneficial for these students. As letter name knowledge facilitates letter sound knowledge, pairing letter name and sound instruction together makes sense, particularly for letters that have their sounds in their names. The order of an alphabet scope and sequence can be further determined by collectively considering factors mentioned above that facilitate alphabet learning, such as letter frequency, uppercase and lowercase letter similarity, and distinctness of visual features. Findings from the meta-analysis by Piasta and Wagner (2010) add support to conducting early literacy intervention for at-risk students in small groups, for longer durations, and in an explicit manner. Extending

alphabet instruction into reading and spelling instruction may be needed for better transfer of skills.

Components and Organization of Shared-Book Reading

For students who are prereaders, meaning-focused instructional components are often delivered within the context of shared-book reading. Recent syntheses and a number of studies have investigated the effects of shared-book reading on the vocabulary and language skills of students in kindergarten (Beck & McKeown, 2007; Biemiller & Boote, 2006; Coyne, McCoach, & Kapp, 2007; Coyne, McCoach, Loftus, Zipoli, & Kapp, 2009; Coyne, Simmons, Kame'enui, & Stoolmiller, 2004; Gormley & Ruhl, 2005; Justice, Meier, & Walpole, 2005; Penno, Wilkinson, & Moore, 2002; Trivette, Simkus, Dunst, & Hamby, 2012). For the purposes of this review, I was interested in studies involving kindergarten students identified as at risk or with early literacy difficulties and instruction delivered in small groups. Because of the limited number of studies located, studies or syntheses involving preschool students, students characterized as typically-developing, or whole-class or large-group instruction were included.

Trivette et al. (2012) and Gormley and Ruhl (2005) conducted syntheses of shared-book reading studies. Trivette et al. located 16 studies that evaluated the effect of repeated readings of stories or compared the effects of one read versus more than one read. Nine studies involved students characterized as typically developing, seven studies included children at risk, and only one study respectively involved students with developmental disabilities and students who were English language learners. Participants included students who were not yet in kindergarten as well as students who were of kindergarten age. Study outcomes were categorized into the following three categories: story-related vocabulary, story-related comprehension, and expressive language. Results indicated larger effect sizes on outcome measures for reading the

same book at least four times every three days, reading for 20 minutes of more, focusing on one or two books at a time, responding positively to child comments, using manipulatives to encourage child participation, giving answers to child questions, and asking open-ended questions. The average effect size based on age was as follows: 1.60 for students aged 49 to 60 months, 0.84 for students 61 months or older, and 0.49 for students less than 48 months. The average effect sizes for typically-developing students and students at risk were 0.87 and 0.49 respectively.

Gormley and Ruhl (2005) focused their review on seven studies targeting preschool-aged students at risk or with identified disabilities and utilizing techniques of dialogic shared-storybook reading in inclusive settings. For specifics and procedures pertaining to dialogic reading techniques, see Flynn (2011), Morgan and Meier (2008), and Whitehurst (1992). From the six studies with positive outcomes and studies previously reviewed by Karweit and Wasik (as cited in Gormley & Ruhl, 2005), Gormley and Ruhl identified a number of general instructional patterns with research support. Their recommendations included the following activities for preschool-aged students with and without disabilities: questioning and prompting, modeling, evaluative comments, defining, follow-up activities, labeling, and summarizing.

A few studies reviewed were implemented in the context of whole-class or large-group instruction (Beck & McKeown, 2007; Biemiller & Boote, 2006; and Penno et al., 2002). All studies involved reading storybooks aloud and selecting a number of words from each story for vocabulary instruction. Biemiller and Boote (2006) and Penno et al. (2002) primarily explained word meanings during read alouds, whereas Beck and McKeown (2007) provided vocabulary instruction after reading. In one of their studies, Biemiller and Boote included systematic vocabulary reviews and asked comprehension questions following the read aloud. Beck &

McKeown included examples of words in additional contexts and opportunities for students to make judgments about word examples. In the next paragraph, results on outcome measures are discussed.

Biemiller and Boote (2006) and Penno et al. (2002) found that students made greater gains in vocabulary for words whose meanings were explained during read alouds than for words that were not, although in Penno et al. some incidental word learning did occur. Word learning was also enhanced through frequency of exposure, that is, through rereadings of the story. Specifically, Biemiller and Boote found that the kindergarten students in their study made 23% gains on word meanings from four readings as opposed to 16% gains from two readings. Biemiller and Boote further found that gains increased from the following changes: increasing the number of words selected for explanation from four to six to seven to nine, including daily reviews of word meanings, and including a comprehensive reviews on the final day of a five-day instructional cycle. In their study, Penno et al. found that higher ability students made larger gains.

In their study, Beck and McKeown (2007) found that significantly more word learning occurred for kindergarten and first grade students who participated in after reading activities than students in the control condition. Examples of activities include the following: contextualizing words by referring back to the sentence in the story in which words occurred, providing word meanings, having children repeat the words, giving examples in additional contexts, having children make judgments about word examples, and having children create their own examples. In a second study, Beck and McKeown compared word learning for words taught in one session as compared to words taught in three sessions. On average, words in the former condition were taught for about 6.6 minutes and five encounters, and words in the latter condition were taught

for about 27.6 minutes and 20 encounters. Results indicated that students learned significantly more in the latter condition.

Some experimental studies were conducted in the context of small-group interventions with kindergarten students identified as at risk based on measures of vocabulary, letter naming, or phonological awareness (Coyne et al., 2007; Coyne et al., 2009; Coyne et al., 2004; Justice et al., 2005). All studies included direct vocabulary instruction within the context of shared-book reading interventions. Across all studies, words selected from storybooks were either chosen for incidental exposure, taught briefly during read alouds in what was termed either embedded or elaborated vocabulary instruction, or taught more extensively during and after read alouds for extended vocabulary instruction. The studies reported in Coyne et al. (2007) are described in detail next to provide a description of incidental learning and embedded and extended instructional conditions. This description is then followed by a synopsis of all four studies.

The purpose of the two studies reported in Coyne et al. (2007) was to investigate the effectiveness of direct vocabulary instruction in small-group interventions in kindergarten. The first study compared "extended instruction of target vocabulary to incidental exposure during story reading" (p. 76). The second study compared extended instruction to embedded instruction. In study one, students were taught in groups of three to four students in 20-30 minute sessions. Six target words were chosen for the story; three were taught incidentally only through three readings of the story and three were taught in a direct and extended manner. Prior to reading the story, target words for extended instruction were spoken by the interventionists and repeated by students. During the reading, students raised their hands when the heard a target word. Interventionists reread the sentence and provided a simple definition of the word. The sentence was again reread, inserting the simple definition into the sentence. Students again

stated aloud the target word. The story reading took approximately 10-20 minutes. After reading, students participated in activities to interact with words in rich and varied contexts. Words were first reintroduced by referring back to how the word was used in the story. Post-reading activities included recognizing examples of target words, answering questions about target words, formulating sentences with target words, and responding to sentences that contained multiple target words. Open-ended questions were used to extend student answers when necessary, and corrective feedback was also given. Results indicated that words taught during extended instruction were learned to a greater extent than those encountered incidentally.

In study two, the same procedure was followed, except three words were taught using extended instruction and three words were taught using embedded instruction. In the embedded condition, students were given simple definitions of words during the read aloud. Sentences were then reread with the target word being replaced by the simple definition. Results indicated that words taught during extended instruction were learned to a greater extent than those taught in the embedded condition. In the next two paragraphs, results of all four shared-book reading intervention studies are discussed collectively.

Justice et al. (2005) and Coyne et al. (2007) found no evidence of word learning through incidental exposure (i.e. hearing the words during story read-alouds) for students at risk. All studies found evidence that direct instruction of vocabulary words positively impacted students' word learning. Whereas Justice et al. found that children with lower initial vocabulary scores made greater gains on elaborated words than children with higher initial vocabulary scores, Coyne et al. (2004) found that students made similar gains regardless of initial vocabulary knowledge. However, in comparison to students who did not receive the intervention, students in the Coyne et al. study with lower initial receptive vocabulary scores made the greatest gains in

taught vocabulary. Coyne et al. (2007) and Coyne et al. (2009) both found that word learning occurred in both the embedded and extended vocabulary conditions. However, results on outcome measures indicated that students' word learning was more complete for words taught in the extended vocabulary conditions.

Studies in this section inform the teaching of meaning-focused components for kindergarten students in a number of respects. First, rereading of storybooks has led to enhanced vocabulary, expressive language, and story-related comprehension outcomes. Second, certain types of interactive techniques appear to have enhanced student learning, such as questioning and prompting, summarizing, asking open-ended questions, and responding to student questions. Third, explicit teaching of word meanings has enhanced the word learning of kindergarten students, including those at risk. Although some evidence exists that students do learn words incidentally from simply listening to stories (Penno et al., 2002), evidence from intervention studies provides contradictory evidence for at-risk populations (Justice et al., 2005; and Coyne et al., 2007). As the number of words that can be taught is limited, studies from Coyne et al. (2007) and Coyne et al. (2009) provide guidelines for choosing words for instruction and suggestions for teaching some words briefly and others more in depth.

Thus far, I have reviewed studies that inform the literature base pertaining to early literacy interventions for kindergarten students at risk for future reading failure. Additionally, I have reviewed studies that specifically inform the literature pertaining to effective components of code-focused and meaning-focused components. For code-focused components, alphabet knowledge instruction was specifically targeted as effect sizes tend to be lower for alphabet knowledge than for phonemic awareness. Now, I turn my attention to studies that inform the literature as to the best predictors of kindergarten response to intervention.

Predictors of Response to Intervention

The goal of early reading intervention research is to boost early literacy foundational skills to levels that will better ensure future reading success for students at risk. Despite receiving additional instruction, some students have demonstrated an inadequate response. If the characteristics of nonresponders and most significant predictors of response to early reading intervention can be identified and better understood, intensive intervention supports can be organized and implemented for these students in a more preventive and proactive manner.

Al Otaiba and Fuchs (2002) conducted a review of literature to investigate the characteristics of nonresponders to early literacy intervention. Of 21 studies that measured phonological awareness skills, 16 studies found this characteristic to be an important correlate to treatment responsiveness. Attention was an important correlate in seven of nine studies, rapid automatic naming in five of seven studies, and phonological memory in four of seven studies. Out of 15 studies reviewed, five reported intelligence quotient (IQ) to have an important relationship to unresponsiveness and seven did not.

Nelson, Benner, and Gonazlez (2003) completed a meta-analytic review that included 19 studies from the Al Otaiba and Fuchs and 11 additional studies that met their inclusion criteria. The authors' primary purpose was to identify the learner characteristics with the largest influence on the treatment effectiveness of early literacy interventions. The biggest predictors in order of magnitude were rapid naming, phonological awareness, problem behavior, alphabetic principle, memory, and IQ. Demographic learner characteristics such as disability/retention, ethnicity, and grade level did not have large influences on treatment effectiveness.

In a later study, Al Otaiba and Fuchs (2006) identified characteristics of nonresponders in subsequent grade levels after kindergarten and first grade intervention. "Nonresponsive students scored about 1.5 standard deviations (SDs) lower than always responsive students on measures of vocabulary, rapid naming, and problem behavior" and 1 SD lower on verbal memory (p. 426).

Other studies have investigated the relationship between measures administered prior to instruction or intervention at the beginning or during the kindergarten year and measures administered post-intervention or in subsequent grades (Catts, Fey, Zhang, & Tomblin, 2001; Catts, Nielsen, Bridges, Liu, & Bontempo, 2013; Hagan-Burke et al., 2011; Ortiz et al., 2012; Parrila, Kirby, McQuarrie, 2009; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004; Simmons et al., 2013; Torgesen et al., 1999; Vellutino et al., 2008). In addition, the National Early Literacy Panel (NELP) (2008) report investigated predictors of children in kindergarten and earlier to later literacy outcomes. With the exception of Torgesen et al. (1999), knowledge of letter names or letter identification was reported to be a significant predictor of later reading outcomes in all of these studies. Schatschneider et al. (2004) noted that knowledge of letter names was a more significant predictor in the beginning of kindergarten and diminished in its relationship over time due to a ceiling effect. In the studies by Catts et al. (2001), Catts et al. (2013), and Ortiz et al. (2012), measures of letter naming, letter identification, or letter-word reading were found to be the best overall predictors. Catts et al. (2013) further found that growth in letter naming fluency (LNF) during the first six weeks of the school year did not add to the prediction of later reading outcomes, but growth in the first half of the school year did.

Measures of phonological awareness were also reported to be significant predictors in all studies with the exception of Hagan-Burke et al. (2011). Simmons et al. (2013), who used a measure of sound matching, suggested augmenting that measure with a measure sound blending

for future studies. Catts et al. (2013) noted that a measure of sound matching was one of the strongest predictors and was a better predictor than DIBELS (2002) Initial Sound Fluency (ISF). The First Sound Fluency subtest, which has replaced ISF in the most recent version of DIBELS, was not used in this study. In their study, Parrila et al. (2009) found that phonological awareness, as measured by sound isolation and blending phonemes, was a better overall predictor of first and second grade reading outcomes than a naming speed task or a letter identification task.

A number of studies found rapid automatic naming (RAN) to be a significant predictor of later reading outcomes (Catts et al., 2001; Catts et al., 2013; Hagan-Burke et al.; NELP, 2008; Parrila et al., 2009; Schatschneider et al., 2004; Simmons et al., 2013; Torgesen et al., 1999). Schatschneider et al. (2004) and Vellutino et al. (2008) found letter sound knowledge to be a significant predictor. Difference in gain scores in letter sound knowledge from December to March of kindergarten was one of the largest predictors reported by Vellutino et al. In the Schatschneider et al. study, knowledge of letter names was consistently more predictive than letter sound knowledge in the beginning of kindergarten, but not significantly so. In addition, Schatschneider et al. found rapid automatic naming of letter symbols to be a better predictor of reading fluency than rapid automatic naming of objects. Interestingly, measures of phonological awareness, letter knowledge, and rapid automatic naming were better predictors of reading comprehension than measures of oral language. Schatschneider et al. suggested that a more comprehensive battery of oral language measures may produce different outcomes.

Many of the studies that investigated the best predictors of future outcomes for kindergarten students also examined what combination of subtests explained the most variance in later outcomes (Catts et al., 2013; Parrila et al., 2009; Schatschneider et al., 2004; Vellutino et

al., 2008). For example, Schatschneider et al. (2004) found that measures of phonological awareness, knowledge of letter sounds, and rapid naming of letters accounted for the most unique variance in reading outcomes. Vellutino et al. (2008) used the following measures at the beginning of kindergarten to predict later response to intervention: letter identification, number identification, alliteration, rhyme detection, and rapid object naming. With this model, only rhyme detection and number identification contributed unique variance and the overall classification accuracy was rated by the authors as fair. Parrila et al. (2009) found that phonological awareness and naming speed in kindergarten "accounted for large unique variance in all reading measures" given in first, second, and third grade (p.15). With phonological awareness and naming speed in their predictive model, letter recognition, short term memory, and speech articulation rate added little to the predictive power of the models. More recently, Catts et al. (2013) reported that a screening battery of the (LNF) subtest from DIBELS, a sound matching task from the Comprehensive Test of Phonological Processing (CTOPP), and the Rapid Naming of Objects subtest from the CTOPP accurately identified good and poor readers at the end of first grade. In the NELP synthesis, the most consistent predictors of later literacy outcomes were alphabet knowledge, phonological awareness, rapid automatic naming of objects and colors, rapid automatic naming of letters and digits, writing or name writing, and phonological short term memory.

Considered collectively, students investigating the best early predictors of later literacy outcomes paint a rather complex picture. Alphabet knowledge, phonological awareness, and rapid automatic naming appeared to be the most consistently reported predictors. In their recent study, Simmons et al. (2013) reported letter identification, rapid naming, and sound matching to be significant predictors of later literacy outcomes. However, only letter identification was a

significant predictor of word identification, oral reading fluency, and passage comprehension both for students who had received a packaged intervention program and for students who had received school-based interventions. In the four studies reviewed that explored the predictiveness of combinations of subtests, only phonological awareness was a part of all test batteries (Catts et al., 2013; Parrila et al., 2009; Schatschneider et al., 2004; Vellutino et al., 2008). Rapid naming is a significant predictor in three of the four studies. Although alphabet knowledge is reported as a consistent predictor in a majority of the studies and syntheses reviewed, alphabet measures used in Parrila et al. (2009) and Vellutino et al. (2008) did not add to the prediction models.

Summary

In this chapter, I reviewed two main bodies of the literature. The first body of literature consists of experimental studies investigating the effect of early literacy interventions on the early literacy and reading outcomes of kindergarten students at risk for reading difficulties. The second body of literature consists of quantitative studies investigating the predictors and characteristics of responders and nonresponders to early literacy intervention. Studies from the first body demonstrate that supplemental, small-group instruction of early literacy skills is effective for most at-risk students. They also demonstrate that some students do not adequately respond to these support services. The studies reviewed also varied in terms of their intensity. Interventions can be intensified by reducing the size of intervention groups, by using highly-specified instruction, and by increasing the number of sessions per week, the length of sessions, and the number of weeks of intervention. Few studies have investigated the effects of a code- and meaning-focused intervention beginning early in the kindergarten year, implemented daily

for 25-30 minutes in groups of two to three students. Additional research is needed to investigate the effects of more intensive early literacy interventions for the most at-risk students.

Studies from the second body of literature indicated that phonological awareness, letter knowledge, and rapid automatic naming are some of the most significant predictors of early later literacy outcomes. Findings from these studies can be used in the identification process of students at risk for future reading failure. The majority of kindergarten intervention studies reviewed identified students with an initial screening task only. Generally, all students who met the at-risk criteria received the same intervention supports. Put another way, no distinction was made among students with more and less intensive needs at kindergarten entry. In Torgesen et al. (1999), researchers used a two-stage screening process to select students for intervention. In the first stage, students were screened with a letter naming task. In the second stage, students who scored in the lowest 30% on the letter naming task were tested further. Students with the lowest combined scores on a phoneme deletion task and a serial naming of numbers task were selected for the intervention. A similar two-stage process can be used with a slightly larger battery of the most predictive measures of future reading success Students with the overall lowest performance can then be streamlined to receive intensive intervention supports immediately.

CHAPTER THREE: METHODOLOGY

A review of the literature has identified a lack of research measuring the effectiveness of intensive reading interventions on the early literacy outcomes of kindergarten students with the greatest early literacy needs. According to Horner et al. (2005), single-case methodology can be used to establish evidence-based practices. In single-subject studies, all students selected for participation can receive the experimental treatment. As such, single-case designs are particularly useful when ethical concerns arise from withholding instruction from a comparison group in an experimental design (Gay, Mills, & Airasian, 2012). Further, single-case designs are compatible with investigating individual responses to manipulation of independent variables (Barlow et al., 2009). With this type of design, the individual participant is the unit of analysis, where "each participant serves as his or her own control" (Horner et al., 2005, p. 166). Consequently, I employed a single-case approach for this study to investigate the individual responses of the most at-risk students to an intensive early literacy intervention.

Research Design

Single-case research is a type of experimental research. Single-case designs are appropriate when the research question is examining the causal relationship between independent and dependent variables and evaluating the performance of a specific individual under certain conditions (Horner et al., 2005). Kratochwill et al. (2010) noted the overarching goal of single-case research is to determine whether an intervention is more effective than the current "baseline." A number of design types exist for a single-case approach.

For this study, I utilized a multiple baseline design across participants as described in Barlow et al. (2009). Multiple baseline designs are appropriate when the effects of the independent variable cannot be reversed once treatment is withdrawn (Gay et al., 2012), as may

be the case when learning new academic knowledge. Multiple baseline studies begin with a baseline phase involving all participants. Experimental control in a multiple baseline across participants design is achieved through staggered introduction of the independent variable. Barlow et al. and Horner et al. (2005) recommended a minimum of three baselines, or phase repetitions. In their panel report for the What Works Clearinghouse, Kratochwill et al. (2010) specified a minimum total of six phases (i.e. three baseline and three intervention phases), with at least five data points per phase. They further specified that three demonstrations of an effect are necessary to provide evidence of a causal relation between the independent and dependent variable. Without three demonstrations, the study provides no evidence of a causal relation. With three demonstrations of an effect and no demonstrations of non-effect, the study provides strong evidence of a causal relation. With three demonstrations of an effect and one demonstration of a non-effect, the study provides moderate evidence of a causal relation.

To illustrate, as the first participant or group is exposed to the independent variable, at least two other participants or groups remain in the baseline phase. A cause and effect relationship is established by the following two occurrences: 1) a change of rate appears in the dependent variable after application of the independent variable for participants receiving the treatment and 2) the rate of behaviors remains relatively constant for participants still in the baseline phase (Barlow et al., 2009). At this point, the independent variable is introduced to the next participant or group, with at least one participant or group remaining in baseline. This same process is then repeated at least one more time. According to Barlow et al. (2009), the temporal sequencing element of multiple baseline studies is vital to ruling out other factors that could account for an observed change in behavior.

According to Horner et al. (2005), a multiple-baseline across participants design controls for threats to internal validity because of inherent within- and between-participant comparisons, whereas threats to external validity require study replication across participants, settings, or materials. Consequently, detailed descriptions of the independent variable, dependent variable, baseline condition, participants, setting, and the process of participant selection are necessary (Horner et al.; Wolery & Ezell, 1993). Single-subject designs further require specific features of dependent and independent variables. Dependent variables need to be: operationally defined using specific language, measured repeatedly across phases, assessed frequently for inter-assessor agreement, and selected for their social significance. Independent variables also require operational definitions with specific language. The independent variable is actively manipulated and monitored for implementation fidelity (Horner, et al., 2005).

Research Questions

In this study, I sought to answer three questions: 1) Is an intensive early literacy intervention effective in improving the performance level and rate of improvement of letter name knowledge of most at-risk kindergarten students beyond baseline levels? 2) Is an intensive early literacy intervention effective in improving the performance level and rate of improvement of letter sound knowledge of most at-risk kindergarten students beyond baseline levels? 3) Is an intensive early literacy intervention effective in improving the performance level and rate of improvement of phonemic awareness, as measured by first sound identification, of most at-risk kindergarten students beyond baseline levels?

Participants and Setting

Participants for this study were selected from four kindergarten classrooms from one school building based on results from early literacy screening and diagnostic assessment data.

The students designated with the highest at-risk status were selected to receive an intensive early literacy intervention. The intervention took place in an elementary school in a moderately-sized city in the northwest United States. Instruction was delivered in small groups of three students per group in a pull-out setting in one of the school's resource room classrooms. Students attended a full-day kindergarten program and received approximately 80 minutes per day of core reading instruction, which consisted primarily of activities from Reading Street (Pearson Education, 2011). According to Pearson Publishing, Reading Street includes systematic and explicit instruction in phonemic awareness, phonics, fluency, vocabulary, and comprehension. Additionally, students designated at risk based on the school's universal screening measures received 20-25 minutes of small group instruction (three to seven students) in Reading Mastery (SRA, 2008) in lieu of whole group instruction from Reading Street. Generally, small-group instruction occurred during core instructional time and therefore was not in addition to core instruction. During intervention time, students missed various classroom activities (e.g. free choice, teacher read aloud, whole-group instruction), depending on when the intervention occurred and the child's classroom schedule.

Participant Selection

Participants were initially selected for additional screening based on results from the school-based universal screening process that takes place in September. The school-based screening process for beginning of year kindergarten will include the Letter Naming Fluency (LNF) and First Sound Fluency (FSF) subtests from Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Next (Good & Kaminski, 2011). DIBELS Next is an assessment system consisting of screening and progress monitoring measurement tools. The DIBELS Next measures are indicators of early literacy skills.

The LNF subtest is a one minute, timed test measuring students' early alphabet knowledge, specifically the ability to automatically name letters. The test consists of randomly arranged uppercase and lowercase letters in ten rows of ten letters per row. Students are instructed to say the name of each letter. Students receive one point for each correctly identified letter name in one minute. The DIBELS Next Technical Manual (Good & Kaminski, 2011) reports alternate-form reliability as 0.86. Inter-rater reliability is reported as 0.99. Criterion validity as correlated to the Group Reading Assessment and Diagnostic Evaluation (GRADE) in the beginning of kindergarten is 0.39.

The FSF subtest measures an examinee's ability to identify the first sound or sounds of spoken words and is considered a test of phonemic awareness. It is also a one minute, timed test that consists of three- to five-sound words. Words are stated aloud one at a time by the test administrator. The visible form of the word is not seen by the test taker. The student listens to the word and responds with the first sound of the word. Students receive two points for correctly identifying the first sound of the word and one point for saying the first sound followed by additional sounds in the word (as specified in the test booklet and as long as the entire word is not repeated). The final score is the total number of points earned in one minute. The DIBELS Next Technical Manual reports alternate-form reliability as 0.83 and 0.52 from two different studies. Inter-rater reliability is reported as 0.94. Criterion validity as correlated to the GRADE in the beginning of kindergarten is 0.52.

Measures to Identify Students at Risk

Thirty-eight students were identified as at risk for future reading failure according to DIBELS Next recommendations. Additionally, each of the four kindergarten classroom teachers was given the opportunity to nominate up to two students for the initial screening based on any

assessment results from the core instructional program and teacher observation of the child's literacy learning. No additional students were nominated. One student who was nonverbal was excluded from the study.

Permission forms were sent out to parents and legal guardians for participation in the study for the thirty-seven remaining students. Permission was obtained for twenty-seven students for a consent rate of 73%. Student assent was obtained for twenty-five of the twenty-seven students for an assent rate of 93%. One student did not give assent. A second student was excluded from the study at this time because of a change in placement due to significant difficulties interacting with peers, participating in small and whole group classroom activities, and transitioning from one activity to the next.

The screening consisted of a two-part process. The initial screening took place at the end of October and beginning of November and consisted of the Sound Matching (SM) subtest from the Comprehensive Test of Phonological Processing-2 (Wagner, Torgesen, Rashotte, & Pearson, 2013) and the AIMSweb (Shinn & Germann, 2006) Letter Naming Fluency (LNF) subtest. The SM subtest measures a child's ability to identify words with the same beginning and ending sounds as a given target word. The examiner presents a target word followed by three additional words, of which one matches to the target word. A student's raw score is the number of correctly identified words with the same beginning or ending sound as the target word. In the CTOPP-2 manual, internal consistency for four to six year olds is reported as .93. Test-retest reliability and inter-scorer reliability for four to six year olds is reported as .78 and .96 respectively. Average criterion validity is reported as .54, or of large magnitude.

The AIMSweb LNF subtest consists of the same basic procedures and scoring as described above for the DIBELS Next LNF subtest. AIMSweb is an assessment system consisting of screening and progress monitoring measurement tools. The AIMSweb LNF test was used in lieu of DIBELS Next LNF to be consistent with the alphabet knowledge dependent measures used in this study (as discussed in the "Dependent Measures" section later in this chapter). The fall of kindergarten benchmark probe was used in the screening process. The AIMSweb technical manual (Pearson, 2012) reports test-retest, alternate-form, and interscorer agreement as .90, .80, and .94 respectively as administered in the spring of kindergarten. Criterion validity for fall, winter, and spring kindergarten administration correlated to spring of third grade administration of the Illinois Standards Achievement Test and Minnesota Comprehensive Assessment range from .51 to .55 and .53 to .60 respectively. Criterion validity for spring administration of LNF and the Woodcock-Johnson Revised (Woodcock & Johnson, 1990) Broad Reading and Reading Skills Composites are reported as .63 and .75 respectively.

Following the procedure of Scanlon et al. (2005), all students who scored below the 30th percentile on either assessment were considered to be at risk. A larger battery of assessments was then administered to each student, similar to the process used in Scanlon et al. and Torgesen et al. (1999). Additionally, teachers were again given the opportunity to nominate students for the larger battery.

Twenty-one of the 25 students, or 84%, were identified as at risk based on the initial screening. Of the four students designated as "not at risk," one was nominated by his classroom teacher to participate in the larger battery of assessments. Study participation for the other three students ended at this point. The larger assessment battery consisted of measures of phonological awareness, alphabet knowledge, and rapid automatic naming and was administered

in November. These categories and specific measures were chosen based on a review of the literature of previous research investigating the most significant predictors of future reading outcomes as described in the Review of the Literature.

Phonological Awareness

Phonological awareness measures consisted of the Blending Words (BW) and Elision (EL) subtests of the Comprehensive Test of Phonological Processing-2 (CTOPP-2). The CTOPP-2 measures reading-related phonological processing skills. The BW and EL subtests combine with the SM subtest used in initial screening to create a composite score for kindergarten-aged examinees. The BW subtest measures a child's ability to blend sounds to form words. Each item is presented in sound segments, at the syllable, onset-rime, or phoneme level. The raw score is the number of correctly blended words. In the CTOPP-2 manual, internal consistency is reported as .86, or at a desirable level. Test-retest reliability and inter-scorer reliability is reported as .75 and .97 respectively. Average criterion validity is reported as .68, or of large magnitude.

The Elision subtest measures a child's ability to delete sound segments from spoken words. In this subtest, removal of the specified sound segment from the spoken word results in a new, real word. The raw score is the number of correctly identified words. In the CTOPP-2 manual, internal consistency is reported as .91, or at a desirable level. Test-retest reliability and inter-scorer reliability is reported as .82 and .96 respectively. Average criterion validity is reported as .70, or of large magnitude.

Alphabet Knowledge

Alphabet knowledge was measured with the kindergarten Phonological Awareness Literacy Screening (PALS-K; Invernizzi, Juel, Swank, & Meier, 2011) Alphabet Knowledge (AK) and Letter Sounds (LS) subtests. Both subtests are untimed. The AK subtest consists of all 26 lowercase letters presented in random order on one page. The student's score is the number of correctly identified letters. The LS subtest consists of 23 uppercase letters and three digraphs (i.e. sh, ch, and th). The student's score is the number of correctly identified letter sounds. As reported in the PALS-K Technical Reference manual, retest reliability correlations for the AK and LS subtests were found to be 0.92 and 0.88 respectively. Criterion validity for each subtest is not reported. The correlation between the Fall PALS-K summed score and the spring Stanford-9 score was found to be 0.70.

Rapid Automatic Naming

Rapid automatic naming (RAN) was measured by the Rapid Naming of Objects (RNO) subtest. Examinees name six different common objects randomly and repeatedly arranged on a one-page sheet in four rows of nine letters per row. The student's raw score is the total number of seconds to name all objects on the entire page. In the CTOPP-2 manual, alternate form reliability for five and six year olds is reported as .84 and .91 respectively. Test-retest reliability and inter-scorer reliability for four to six year olds is reported as .86 and .96 respectively. Average criterion validity is reported as .70, or of very large magnitude.

The stage two screening battery took approximately 20-30 minutes to administer to each student. Students were given a short break during testing if needed based on the discretion of the examiner. If a student had difficulty attending, the rest of the battery was administered at a later time.

Following administration of the two-part screening process, raw scores from the screening measures were converted to four *z* scores and summed to create a summed *z* score. A *z* score "is the number of standard deviation units that a score is above or below the mean" (Steinberg, 2011, p. 97). *Z* scores were calculated by first finding the mean of raw scores from each student who participated in both the initial and extended screening battery. Each student's standard deviation from the mean was then calculated and converted to a *z* score. For example, if a student's score was one standard deviation below the mean, the student's *z* score was -1.0. Conversely, if a student's score was one standard deviation above the mean, the student's *z* score was 1.0.

The first *z* score was calculated from the combined raw scores of the three phonological awareness subtests from the CTOPP-2. The combined scores were used based on results from the National Early Literacy Panel report indicating that phonological awareness composite scores were more predictive of future decoding skills than individual subtests, as well as based on the suggestion from Simmons et al. (2013) to augment the CTOPP SM subtest with the BW subtest. The second *z* score was calculated from student raw scores on the LNF subtest. The third *z* score was calculated from the combined raw scores of the PALS-K AK and LS subtests. The fourth *z* score was calculated from student raw scores on the RON subtest from the CTOPP-2. Because low scores on the RON subtest are superior to high scores, the *z* scores for this subtest were then reversed. The four *z* scores were then added together to create a combined *z* score. Students were rank ordered based on their summed *z* scores. The nine students with the lowest summed *z* scores were selected to receive the intervention.

Because two of the *z* scores pertained to alphabet knowledge, this construct was weighed more heavily in the combined *z* score equation than phonological awareness or rapid automatic

naming. Alphabet knowledge was given more weight due to a number of prior studies that indicated that alphabet knowledge is the single best predictor of future reading outcomes for young children. To confirm this procedure, summed *z* scores were calculated by removing each of the alphabet knowledge subtests from the equation to observe any changes in which students were identified as most at risk. When the LNF *z* scores were removed, the first eight students identified by the original equation as most at risk were again identified (although not in the same order). When the PALS *z* scores were removed, the same nine students were identified (although again not in the same order).

During baseline, one of the nine students made steady progress on two of the three dependent variable measures and appeared to be benefiting from the school's instructional planning. At this time, progress monitoring began for the student who was rank ordered number 11, who, according to the school's data and teacher observation, was not responding adequately to the school's instructional programming. This was confirmed through baseline data and this student replaced the student who was making steady progress.

Demographic data for the final nine participants can be seen in Table 1. All student names are pseudonyms. One-third of the participants are female. According to school records, two-thirds of participants were listed as white and one-third as Native American. Eight students qualified for free lunch and one student for reduced lunch. One of the nine students had an individualized education program with speech and language and academic goals. One student had been retained the previous year and therefore was repeating kindergarten. All other students were attending kindergarten for the first time.

Table 1. Participant Demographic Data

Student	Gender	Ethnicity	Lunch Status	Education Program	Year in Kindergarten
Ben	M	N. Amer.	Free	General Ed.	2nd
Chris	M	White	Free	General Ed.	1st
Sam	F	N. Amer.	Free	General Ed.	1st
Jill	F	White	Free	Special Ed.	1st
Robin	F	White	Free	General Ed.	1st
Tyler	M	White	Free	General Ed.	1st
Kevin	M	White	Reduced	General Ed.	1st
Tony	M	N. Amer.	Free	General Ed.	1st
Frank	M	White	Free	General Ed.	1st

Dependent Variables

The dependent variables for this study were letter-name knowledge, letter-sound knowledge, and first sound identification. Letter-name knowledge was operationally defined as the number of letters correctly named in one minute as measured by the AIMSweb Letter Naming Fluency subtest. Letter-sound knowledge was operationally defined as the number of letter sounds correctly identified in one minute as measured by the AIMSweb Letter Sound Fluency (LSF) subtest. First sound identification was operationally defined as the number of correctly identified first sounds in one minute as measured by the DIBELS Next First Sound Fluency (FSF) subtest.

Dependent Measures

During baseline and intervention phases, the progress monitoring probes of the LNF subtest from AIMSweb, the LSF subtest from AIMSweb, and the FSF subtest from DIBELS Next were administered to all participants. The AIMSweb LNF progress-monitoring probes are structured identically to the benchmark probe used in the screening process described previously

in this chapter in the section titled "Measures to Identify Students at Risk." A different probe was administered during each monitoring session.

The LSF subtest is a one minute, timed test measuring students' early alphabet knowledge, specifically the ability to automatically produce the sound of lowercase letters. The test consists of randomly arranged lowercase letters in ten rows of ten letters per row. Students are instructed to say the sound of each letter. Students receive one point for each correctly identified letter sound in one minute. A different probe was administered during each monitoring session.

The AIMSweb technical manual reports test-retest, alternate-form, and interscorer agreement for LSF as .83, .82, and .82 respectively as administered in the spring of kindergarten. Criterion validity for winter and spring kindergarten administration correlated to spring of third grade administration of the Illinois Standards Achievement Test are reported as .43 and .52 respectively. Criterion validity for spring administration of LSF and the Woodcock-Johnson Revised Broad Reading and Reading Skills Composites are reported as .58 and .72 respectively.

As described previously, the FSF subtest is a one minute, timed test that consists of three- to five-sound words. The test administrator states each word aloud and the student is instructed to respond with the first sound of the word. Refer to the "Participant Selection" section of this chapter for a detailed description of this subtest. The FSF progress monitoring probes include 20 different probes. If all twenty probes were administered, I began readministering probes by returning to the initial probes given. At least a three-month window existed between administrations of the same probe.

All three multiple baseline studies reviewed in Chapter Two used Nonsense Word Fluency (NWF) and Phonemic Segmentation Fluency (PSF) subtests as dependent measures (see Gyovai et al., 2009; Musti-Rao & Cartledge, 2007; Samanich, 2003). Typically, these measures are first given to kindergarten students in the middle of the school year. As I intended to begin intervention services earlier in the school year, I chose the LNF, LSF, and FSF tests. The LNF and FSF tests are both recommended to be given in the beginning of kindergarten. The LSF test is recommended to be given in the middle of the school. However, I chose to include this measure for a number of reasons. First, producing the sounds of letters was a significant component of the intervention (as described later in this chapter). Second, the LSF test requires the examinee to convert letter symbols to their corresponding sounds, a subskill of the decoding process. Third, the NWF subtest consists of three-sound nonsense words in which the examinee either says the sound of each letter or blends the sounds together as a word. The LSF subtest consists of a subskill of the NWF subtest and was therefore considered to be a more sensitive and appropriate measure for participants who would likely have very little letter sound knowledge.

Baseline and Intervention Phases

In the baseline phase, Letter Naming Fluency, Letter Sound Fluency, and First Sound Fluency progress monitoring probes were initially administered by the researcher twice per week. The first three students to demonstrate a stable baseline on all three subtests were considered Group A and began receiving intervention. The remaining students continued in the baseline phase. Two criteria were used to determine when the next group of students began intervention and are as follows: 1) when students receiving the intervention demonstrated a change in rate of growth as compared to their own baseline data, and 2) when three remaining students in the baseline phase demonstrate a stable baseline on all three subtests (Kennedy,

2005). The next group to receive the intervention was termed Group B. At this time, the assessment schedule was relaxed for Group A students to once a week. The same procedure was followed to determine when to begin the intervention phase for the last three students, or Group C. Once all students were in the intervention phase, the assessment schedule for all students was extended to once a week. For ethical reasons, the assessment schedule was again extended to once every week and a half due to what I perceived as assessment fatigue.

Students were placed in groups of three students for a number of reasons. First, findings from the synthesis by Slavin et al. (2011) of reading interventions for student who struggle support groups of three students or less for intensive intervention. Second, groups of two to three students are more feasible than one-to-one interventions in terms of resource allocation. I am in part following the suggestion of Lennon and Slesinski (1999) of employing more intensive ratios of instruction for at-risk students with the greatest needs. Third, group sizes of three or less allow for sufficient opportunities for students to respond, as well as for instructors to provide a sufficient amount of corrective feedback (Vaughn et al., 2010). In the intervention phase, each group of three students received intervention for a total of 60 sessions, or approximately 12 weeks. As such, instruction for each of the three groups began and ended at staggered times.

Follow-up

Three and four weeks after the intervention phase, dependent measures were administered to see if students were maintaining their skill levels. Although multiple baseline designs are appropriate when the effects of the independent variable cannot be reversed, it is reasonable to expect that some of the at-risk participants in this study would regress somewhat in their skill level when intensive supports were withdrawn. Vellutino et al. (2006) found that one distinguishing difference between less difficult and more difficult to remediate students was the

ability to maintain their performance levels when they were no longer receiving intervention. It may be that some students still need some degree of additional systematic and explicit supports to maintain their present skill level or continue progressing at a rate comparable to grade-level peers. In addition, continued monitoring of skill level would be important to determining future level of supports for individual students.

Due to time constraints, only students in Groups A and B were administered follow-up probes. The three-to-four-week timeframe was deemed a reasonable amount of time to gage how students were performing with less intensive supports.

Independent Variable

The independent variable for this study was an early literacy intervention for kindergarten students delivered daily, in small groups of three students, for 25 minutes a day, for 60 total sessions per group. The intervention was researcher-designed. Approximately 16 minutes (or two-thirds) of instruction was comprised of code-focused components and eight minutes (or one-third) of meaning-focused components. Code-and meaning-focused components were delivered in either order. The number of minutes devoted to these two main components is based on a synthesis of kindergarten interventions conducted by Al Otaiba et al. (2005), Lennon and Slesinski (1999), Simmons et al. (2007), and Torgesen et al. (1999). Code-focused and meaning-focused components were included to provide a comprehensive literacy approach. Because of the difficulty of measuring meaning-focused components on a frequent basis, only code-focused components were measured during baseline and intervention phases. I delivered instruction for all three groups of students.

Code-focused instruction

Code-focused instruction initially consisted of approximately 10-11 minutes of alphabet knowledge instruction and four to five minutes of phonological awareness instruction. Over time, these components became more and more integrated. As the intervention progressed, focal points of the code-focused component changed according to the schedule specified in detail below. However, the pacing for each group differed slightly from these timelines based on students' response to intervention as measured by progress monitoring data. In this manner, all groups received the same amount of code- and meaning-focused instruction, but some individualization occurred based on student response to instruction.

Alphabet knowledge instruction began with individual letters and progressed to two- and three-sound word reading, including instruction in high frequency words. Phonological awareness instruction began at the phoneme level but followed a developmental progression from first sound isolation, to last sound isolation, to phonemic blending, and finally to phonemic segmentation (see Paulson, 2004 and Pufpaff, 2009). The introduction of new concepts and skills followed a gradual release of responsibility (GRR) model. The general structure of the GRR model for this study consisted of the following three steps implemented in sequential order: 1) teacher modeling (I do), 2) guided practice (we do), and 3) independent practice (you do).

Letter knowledge instruction consisted of two letters introduced at a time in three-day cycles beginning with uppercase letters. The names and sounds of each letter were taught together using specific and consistent language. As each letter was introduced, its orthographic features were also described. If students had difficulty articulating the sound or had difficulty learning a particular sound, the manner and place of formation was described. Students then received explicit instruction on letter formation. Proper formation was verbalized and modeled.

Students then wrote letters on lined paper. Letters were first traced and then written independently. When needed, scaffolding was provided by providing starting points and creating additional dashed letters to trace.

After the first two uppercase letters were introduced, lowercase letters were introduced in the same manner in the next lesson. Following one more day of additional practice with both uppercase and lowercase letters, two new letters were introduced. With this schedule, students were exposed to each letter of the alphabet within an eight-week timeframe. Each lesson began with a review of previously learned letters and an instructional focus on letters students were having the most difficulty with. Reviews consisted of students identifying both the name and sound of a given letter symbol. As an instructional support, students were given three-dimensional plastic uppercase and lowercase letters to manipulate, feel, and describe. In some sessions, students found newly- and previously-learned letters in books or played an alphabet game.

As the participants targeted for this study had limited alphabet knowledge, the scope and sequence of letter names and sounds was carefully organized according to a number of factors. These factors include: whether or not the sound of the letter is in its name, the frequency of letter in print, the utility of the letters in creating consonant-vowel-consonant words, the structural features of the letter, the similarity between uppercase and lowercase letters, and manner and place of letter sound articulation. The scope and sequence was as follows: T, P, t, p, K, N, k, n, F, I, f, i, B, L, b, l, C, M, c, m, S, A, s, a, D, J, d, j, O, R, o, r, G, V, g, v, H, Z, h, z, W, E, w, e, U, X, u, x, Q, Y, q, y. One vowel letter was introduced early in the sequence so students would be familiar with at least one vowel letter and sound when phonemic blending instruction began.

Phonological awareness instruction consisted of fewer types of phoneme manipulations based on findings from the NRP (2000). Initially, students were taught first sound isolation followed by last sound isolation. The phonemic skills of blending and segmenting were also taught thereafter. As phonemic blending leads naturally to decoding and phonemic segmenting leads naturally to spelling, both of these skills were included. Including both skills is supported in the NELP (2008) finding that only when both skills were taught were moderate effect sizes found on both word reading and spelling outcomes.

Words used in the phonological awareness portion of the lesson included consonant sounds previously taught to reinforce newly taught sounds. Letter sounds that students had difficulty with were integrated in for more frequent, distributed practice. Letter symbols were integrated into the phonemic awareness activities gradually based in part from the technique described by Oudean (2003) and consistent with findings from the NRP. After one week of practice in first and last isolation and once a lowercase letter had been introduced and practiced in at least three sessions, students matched isolated sounds to letter symbols. For example, after students practiced the letter "t" in at least three sessions, they pointed to the letter "t" after correctly isolating the /t/ sound from a spoken word. Letter symbols that students had difficulty with were integrated in more frequently and so differed slightly across groups. This allowed for another element of individualization without changing the main focus of instruction.

During week five, first and last sound isolation instruction was phased out and instruction in phonemic blending of two- and three-sound words with short vowels began. Initially, colored magnets were used to represent each sound of the word. I stated each sound of the word slowly while pointing to different colored magnets positioned in a row. Students then blended the sounds together quickly and said the whole word. At times, each student was given their own

magnet board so they could point to and manipulate the colored magnets. Words chosen for this portion of the lesson consisted mostly of words with the same rime, or from the same word family. Examples of word families used included –it, -in-ap, -at, -og, -op, and –un. The –it word family consisted of words such as sit, pit, hit, lit, fit, etc. Rime families were used to provide an additional scaffold to support student transition to word reading with letter symbols, or blending. Word rimes also expose students to common word patterns and allow for more targeted, repetitious practice.

During the second week of instruction in phonemic blending, magnets with lowercase letter symbols that have been previously taught in at least three sessions replaced the colored magnets. The colored magnets continued to represent letter sounds that have not been taught in at least three previous sessions (as described by Oudean, 2003).

At approximately week eight, the use of colored magnets for phonemic blending was completely phased out. Words were represented using letter tiles or written out using white boards and markers. Words chosen for instruction still consisted of words from the same word family. At approximately week nine or ten (depending on proficiency of student response), word chains replaced word lists with the same rime family. A word chain is a list of words of that has one sound change from one word to the next. An example of a word chain is as follows: sit, sat, pat, pan, pin, tin, tip, etc. This transition from word families to word chains added an element of difficulty but still provided a scaffold of support with repetitious practice.

During week seven, instruction began in phonemic segmentation. Following phonemic blending practice, students practiced segmenting the same words used in the blending activity (as described in Oudean, 2003). Segmenting was initially modeled using different colored magnets

on a small magnetic white board. In some sessions, each student was given her or his own set of different colored magnets. I stated the word, students repeated the word, and then segmented each sound as they placed one magnet in front of them for each sound in the word. Students then pushed the magnets together as they blended the segmented word back together. After two weeks of instruction in phonemic segmenting with manipulatives, lowercase letter symbols replaced the colored magnets or words were written out using white boards and markers.

During week nine, code-focused instruction included word reading and spelling practice of high-frequency words. At this point, all uppercase and lowercase letters had been taught so time spent for letter name and sound instruction was now allocated to high frequency word instruction. However, each code-focused portion of sessions still began with a review of letter knowledge. Fifteen words selected from a high-frequency word list were taught at a pace of three words per week and consisted of the following words: the, a, and, to, you, of, is, he, she, that, I, with, and are. Words chosen for instruction were based on two factors, word frequency and words that corresponded with students' core instruction. Only the word "and" was not specifically targeted during core instruction. Each word was written on an index card. As many high frequency words have irregular patterns that make systematic sounding out problematic, irregular patterns were written using a different colored marker. Attention was brought to each letter of the word by having students say aloud each letter before reading it. An additional activity involved covering up the word and asking questions such as, "What is the first letter of the word," "What is the middle letter," or asking to students to spell the word aloud or on white boards. All sessions included a review of previously taught words. Words that students were having difficulty with were practiced more frequently.

Meaning focused-instruction

Meaning-focused instruction consisted of approximately eight to nine minutes of shared-book reading activities. Books for shared reading consisted of fictional stories and generally included rich and varied language and syntax, as well as vivid illustrations. Books chosen had a limited amount of text due to time constraints. Six vocabulary words were chosen per book for instruction. Three words were designated for extended instruction and three for embedded instruction, as described in the next four paragraphs. See Appendix A for a list of books and vocabulary words chosen for instruction. Additionally, strategies from dialogic reading, as described by Flynn (2011), Morgan and Meier (2008), and Whitehurst (1992), were used to systematically expose students to a variety of question types, teach story structure, and provide opportunities for students to retell stories, also as described in the next four paragraphs.

Stories were repeatedly read in four-day cycles. On day one, stories were read aloud. When a vocabulary word chosen for extended instruction was encountered in text, the word was restated aloud and the sentence or a portion of the sentence containing the word was reread. Students were asked to repeat the word aloud. A simple definition of the word was given and another sentence using the word was stated. At the end of the story, students were asked recall questions pertaining to story characters and main events.

On day two, the story was again read aloud. During the reading, the three words chosen for embedded instruction were discussed in the same manner as words from day one. Following the read aloud, questions were asked pertaining to the characters and plot to lead students to a summary of the story, using illustrations as a scaffold. Prompts were given to encourage student responses, and responses were expanded upon when necessary.

On the third day, students retold the story in their own words using illustrations and open-ended prompts such as "Tell me what's happening on this page." Additional prompting was given and answers were expanded upon when needed. Students were encouraged to incorporate vocabulary words into their retelling.

On day four, words chosen for extended instruction were reintroduced by restating the student friendly definition, giving additional examples of word use, and at times referring back to the context of the story. Students participated in additional activities such as identifying examples and nonexamples of target words, answering questions about target words, or stating their own examples. If time allowed, the story was read a fourth time.

Additional Measures to Determine Response to Intervention

Although the LNF and LSF subtests demonstrate student improvement in alphabet knowledge, they do not provide specific information as to which letter names and sounds students know and still need to learn or practice. As such, the PALS-K Alphabet Knowledge and Letter Sound quick check measures were given approximately once every twelve sessions to each participant to more systematically track which letters names and sounds students know and still need to learn. This information was used to select letters and sounds for review for each group. The Alphabet Knowledge and Letter Sound quick checks are a progress-monitoring tool and consist of the same items as the PALS-K Alphabet Knowledge and Letter Sounds subtests used in the extended screening battery described previously in this chapter in the section titled "Measures to Identify Students at Risk." For each quick check, the order of the items has been rearranged.

Interventionist

I delivered all instruction for each group for all instructional sessions and administered all dependent measures. I have eight years of experience implementing reading interventions to students in kindergarten through eighth grade in public school settings. In addition, I have familiarity with a number of early reading intervention programs and have received training or professional development in administering DIBELS and AIMSweb measures, as well as achievement and diagnostic testing.

Research Assistants

Two graduate students served as research assistants. The research assistants performed three main tasks throughout the study. First, they administered measures during the two-stage screening process. Each research assistant was trained in the administration of each measure. For each measure, I developed a checklist consisting of all major administration and scoring procedures. Research assistants were required to demonstrate 90% or above proficiency on each measure in three consecutive trials prior to administration to study participants. Second, they conducted treatment fidelity checks to document the accuracy of implementation of the independent variable as described below in the section titled "Fidelity of Implementation." Third, the research assistants collected interrater reliability data for administration of the dependent measures as described in the next section.

Inter-Assessor Agreement

As accuracy of scoring for assessment measures is important to interpreting results, inter-assessor agreement was calculated and reported for screening assessments and dependent measures. I observed administration of 27% of screening measures. During these sessions, I collected data alongside the research assistant conducting the assessment. Thirty-three percent

of progress monitoring sessions of dependent measures (i.e. LNF, LSF, and FSF) included one of the research assistants as an additional assessor. Observations occurred in both phases. Kratochwill et al. (2010) recommended inter-assessor agreement on a minimum of 20% of monitoring sessions in each phase. Observations were interspersed throughout both phases. Both the administrator and observer scored each probe during administration. Inter-assessor agreement between the administrator and the observer for both time periods was calculated using interval agreement, or total agreement using the following formula: Agreements / (Agreements + Disagreements) X 100% (Kennedy, 2005).

Fidelity of Implementation

As a measure of how accurately the independent variable was implemented as designed, a treatment fidelity checklist was used to document fidelity of intervention implementation. The checklist consisted of 15 items that corresponded to the main components of the intervention (see Appendix B). Fidelity data was collected for twenty-five percent of sessions for each group. Fidelity observations were interspersed throughout the entire 60-session timeframe. One of the two research assistants attended the entire instructional session and marked each component as either *present*, *not present*, or *not applicable for this lesson*. Components were only marked *not applicable for this lesson* if they were not part of the intervention design for that particular session.

Treatment fidelity was calculated for each session as the percent of components present out of the total number of components applicable for each lesson. At the conclusion of the study, the total percentage of components included was calculated for each group by summing the number of components present from each session and dividing by the total number of components that should have been present. An overall fidelity percentage across groups was

then calculated by summing the number of components present from every session and dividing by the total number of components that should have been present.

Data Analysis

Data from dependent measures were graphed for all participants. Data were analyzed through systematic visual comparison of student scores during the baseline and intervention phases of the study, which is the primary means of examining data in single-case research according to Kennedy (2005). Data were analyzed both within and between phases for each participant for each of the three dependent variables (i.e. LNF, LSF, and FSF).

Within phases, systematic visual comparison consists of three dimensions: the level, trend, and variability of performance (Horner et al., 2005; Kennedy, 2005). The level of performance refers to the mean scores of individual participants within a phase (Kratochwill et al., 2010). The trend of performance refers to the rate of increase or decrease in student scores within a phase, or the slope of the best-fitting straight line (Kratochwill et al, 2010). Trend lines for baseline and intervention phases were created using the Split Middle (SM) technique (Kennedy, 2005). With the SM technique, trend lines are created by first dividing the number of data points in each phase in half. The median score is then found for each half of the data. On the graph, the intersection of the median score and the median number of monitoring sessions for each half of the data is located. A straight line is then drawn between the two median scores. The baseline trend line can then be extended into the intervention phase for a visual comparison of the trend lines for both phases.

Variability refers to the degree to which data points fluctuate around the mean and slope within a phase (Horner et al., 2005). Kennedy (2005) and Kratochwill et al. (2010) explain

variability as the range of data around the slope of the best fitting line. If a high degree of variability exists in baseline, the baseline phase should be extended to establish stability (Kennedy, 2005; Kratochwill et al., 2010).

Between phases, the immediacy of effect and degree of non-overlapping data was analyzed. Kratochwill et al. (2010) defined immediacy of effect as "the change in level between the last three data points in one phase and the first three data points of the next" (p. 18). Generally, the functional relationship between the independent and dependent variable is more convincing with more rapid immediacy of effects (Kennedy, 2005; Kratochwill et al., 2010). In studies with a predicted delayed effect, Kratochwill et al. suggest extending phase length.

The degree of non-overlapping data is an indicator of performance differences between phases in single subject research. Non-overlapping data refers to what percent of data in adjacent phases do not overlap (Parker & Vannest, 2009). The effect of the independent variable on the dependent variable(s) is more convincing with a smaller proportion of overlapping data points (Kratochwill et al., 2010).

A number of non-overlap techniques have been designed to quantify single-subject design intervention effect size. Traditionally, percent of non-overlapping data (PND) has been a commonly used non-overlap technique. However, according to Parker & Vannest (2009), effect sizes from PND can be disproportionately influenced by outliers and the appropriateness of its further use has been questioned (see Wolery et al., 2010 and Maggin, et al., 2011 for more disadvantages). Non-overlap of all pairs (NAP) is a more complete non-overlap technique that individually compares all data points across phases (Parker & Vannest, 2009). In the Parker & Vannest study, NAP was shown to be a better discriminator of single-subject results than PND.

Although NAP did have a ceiling effect and did not discriminate well among the most successful interventions (starting at the 80th percentile), PND did not discriminate among nearly half of the interventions. As compared to PND, NAP was also more highly correlated to R^2, the most commonly used effect size in experimental research. In a more recent study by Parker, Vannest, and Davis (2011), NAP was demonstrated to have greater statistical power than a number of other non-overlap techniques. Insufficient statistical power hinders the ability to reliably identify smaller effects and results in lower precision (i.e. large confidence intervals). Consequently, NAP was used in this study to compare data overlap between phases.

Analysis was enhanced using the conservative dual-criterion (CDC) method as described by Fisher, Kelley, & Lomas (2003) and Swoboda, Kratochwill, and Levin (2010). The CDC method was developed to enable a more systematic, objective, and reliable approach to improving traditional visual analysis (Swoboda et al., 2010). The CDC method stems from the dual-criteria (DC) method in which a line for level (i.e. mean) and a line for trend from baseline are superimposed upon the treatment graph. Adding a line for level was found to reduce Type I errors, or concluding that there was an intervention effect when in fact there was not (Fisher et al., 2003). Two criteria are then used to conclude that a systematic change occurred from baseline phase to treatment phase (Swoboda et al., 2010). Depending on the number of data points in the treatment phase, a specific number of data points in the treatment phase need to be above both the level line and the trend line (specific criteria are given in Chapter Four). The CDC method increases the level and trend lines by 0.25 standard deviations. Fisher et al. found that increasing the two criterion lines by 0.25 standard deviations further reduced the number of Type I errors to tolerable levels.

Summary

A multiple baseline across participants design was employed to investigate the effect of an early literacy intervention on the alphabet knowledge and phonemic awareness of the most at-risk kindergarten students. Nine participants were selected from one school in the western United States through a two-stage screening process designed to identify those kindergarteners most at-risk for future reading failure. The intervention consisted of both code-focused and meaning focused components and was implemented for 25 minutes daily for a total of 60 sessions. During baseline and intervention phases, alphabet knowledge and phonemic awareness skills were measured repeatedly using tests of letter naming fluency, letter sound fluency, and first sound fluency. Throughout data collection, inter-assessor and fidelity observations were conducted on a regular basis. Data was analyzed through systematic comparison of within and between phase patterns and was enhanced using the conservative dual criterion approach.

CHAPTER FOUR: RESULTS

This chapter describes the results of the study. As discussed in Chapter Three, multiple baseline studies require a minimum of three baseline and three intervention phases (Barlow et al., 2009; Horner et al., 2005; Kratochwill et al., 2010). Kratochwill et al. (2010) further specified that three demonstrations of an effect are necessary to provide evidence of a causal relation between the independent and dependent variable. As such, I first present results for all three groups during baseline and intervention phases that document evidence of the causal relation between the independent and dependent variable. In the next three sections, results are discussed for each individual student organized according to the three research questions of the study. Data for all students are described in terms of within-phase and between-phase patterns. Within-phase patterns discussed include the level, trend, and variability of the data. Between-phase patterns include the immediacy of effect and degree of non-overlapping data. Remaining sections present results pertaining to follow-up data, PALS progress monitoring data, inter-assessor agreement for measures given during the screening process, baseline phase, and intervention phase, and fidelity of implementation data for each of the three groups and for all groups combined.

Evidence of a Causal Relation

Figures 1, 2, and 3 display letter naming fluency, letter sound fluency, and first sound fluency scores respectively for all nine students during Baseline and Intervention Phases. After two and half weeks, the independent variable was introduced to three students with relatively stable baseline data on all three dependent measures, who thus became Group A.

Figure 1. Results of Baseline, Intervention, and Follow-up Letter Naming Fluency Data

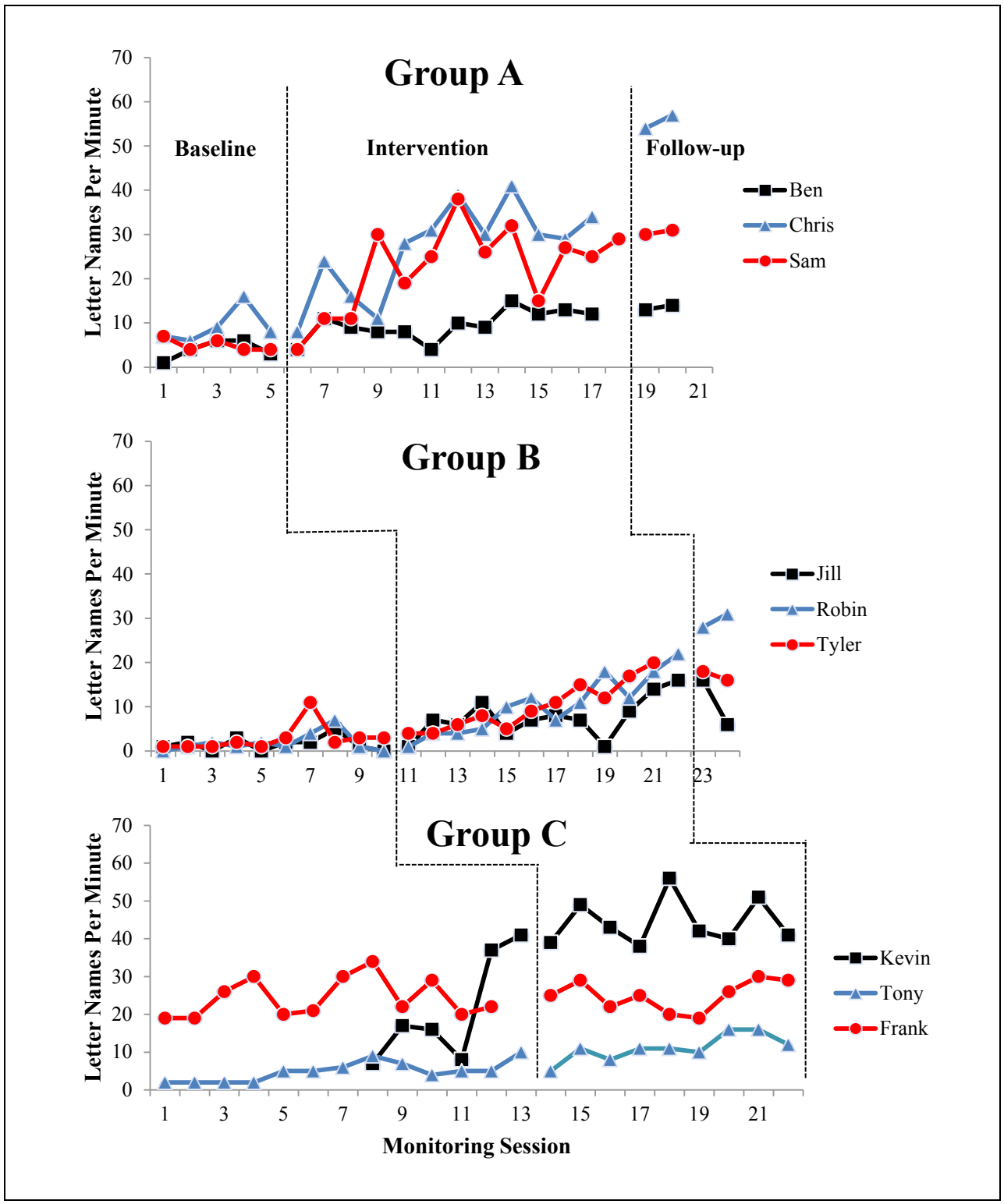


Figure 2. Results of Baseline, Intervention, and Follow-up Letter Sound Fluency Data

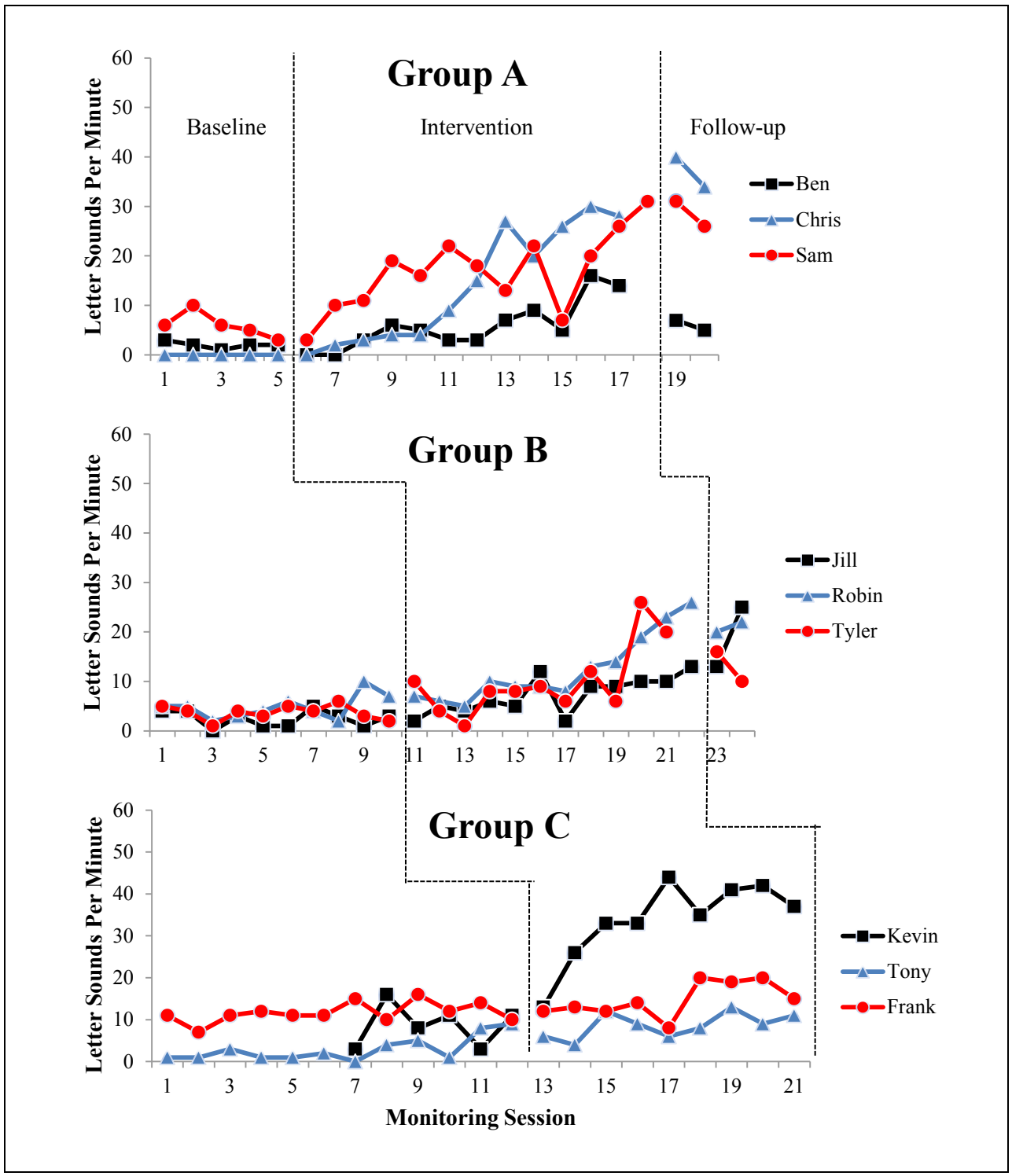
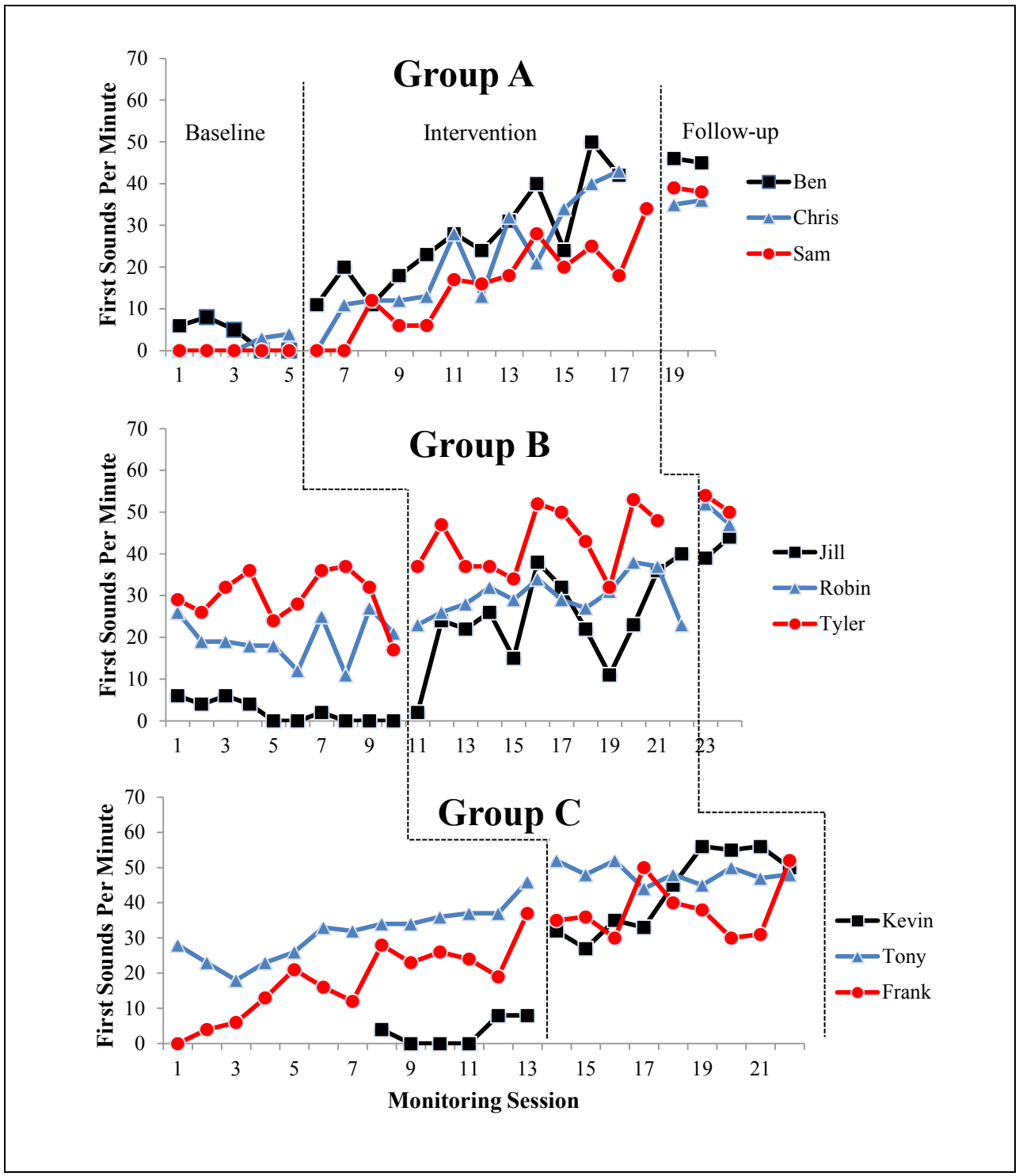


Figure 3. Results of Baseline, Intervention, and Follow-up First Sound Fluency Data



As illustrated, a change of rate occurred for all students in Group A during the first two weeks after instruction began. At the same time, baselines for Jill, Robin, and Tyler remained relatively stable, providing evidence for a causal relation between the independent and dependent variables at time one. Consequently, the independent variable was purposefully introduced to these three students, who became Group B. Noteworthy was a small spike in Robin's LSF data and some variability in both Robin's and Tyler's FSF data. As rates of growth were slight in all three instances, the decision was made to end baseline and begin intervention. Also during this timeframe, baseline testing began for Kevin, who replaced a student who was displaying a steady rate of growth on LNF and LSF, as well as demonstrating the ability to sound out three-sound words in small group instruction.

Increases in skill level for Group B can be observed after three weeks of manipulation of the independent variable. In Figures 1, 2, and 3, the first six data points in the intervention phase for Group B students corresponds to this timeframe. Data for all three students showed steady rates of increase in LNF and LSF. Additionally, FSF data for both Jill and Robin showed increases, including a rather sharp increase for Jill. FSF data for Tyler demonstrated an increase in level, but an increase in rate was inconclusive at this time.

Generally, data for Group C showed the least amount of stability during baseline. All three students had relatively stable baselines on two of the dependent measures but showed growth on one of the dependent measures. Specifically, Kevin's scores on LSF and FSF were stable, but exhibited a sharp rise in LNF performance level just prior to introduction of the independent variable. Tony and Frank's LNF and LSF data showed relative stability, but both students' FSF data was steadily increasing. Because each student had stable baseline data on two of the measures and Group B students were showing increases in rate, the independent variable

was purposefully introduced to Group C. Additionally, two students in Group A, Chris and Sam, were continuing to show steady rates of improvement on all three measures. Ben was not showing growth on either LNF or LSF, but was continuing to grow at a steady rate on FSF. Generally, evidence of a causal relation between the independent and dependent variables existed at time two.

Following introduction of the independent variable for Group C, rates of improvement on the dependent variables were not seen for Tony or Frank. Kevin exhibited a steady rate of improvement on both LSF and FSF. Because steady rates of improvement were generally not apparent, evidence of a causal relation between the independent variable and dependent variables was not present at time three. The next three sections provide a more specific analysis of within and between phase data for each participant according to the three research questions of this study.

Research Question Number One

Is an intensive early literacy intervention effective in improving the performance level and rate of improvement of letter name knowledge of most at-risk kindergarten students beyond baseline levels?

Table 2 displays mean level of performance as well as indicates trend level comparisons for all students in baseline and intervention phases. Mean scores are reported as letter names correct (LNC) per minute. All students experienced an increase in mean scores from baseline to intervention phases. However, Frank's increase was negligible. Mean score differences ranged from 0.7 LNC for Frank to 23.3 LNC for Kevin. Relatively modest increases occurred for Ben,

Jill, Robin, Tyler and Tony. Data for Chris, Sam, and Kevin showed relatively large increases in mean scores from baseline to intervention phase.

Trend of the data was analyzed using the split-middle technique (see Figure 4, 5, and 6). All students' data showed a positive slope in the intervention phase. However, only six of the nine students had a steeper positive slope in the intervention phase than baseline phase. Slopes for Ben, Kevin, and Frank were steeper during baseline. In the trend column in Table 2, students who had more positive slopes in the intervention phase are labeled "increasing." Students who had a more positive slope during baseline are labeled "decreasing."

Table 2. Within Phase Analysis of Letter Naming Fluency Data

Participant	BL Mean	INT Mean	Trend
Ben	4.0	9.6	Decreasing
Chris	9.2	27.1	Increasing
Sam	5.0	22.5	Increasing
Jill	1.6	7.6	Increasing
Robin	1.9	10.3	Increasing
Tyler	2.8	10.1	Increasing
Kevin	21.0	44.3	Decreasing
Tony	4.9	11.1	Increasing
Frank	24.3	25.0	Decreasing

Note. BL = baseline; INT = intervention.

Figure 4. Group A Letter Naming Fluency with Split Middle Technique Trend Lines

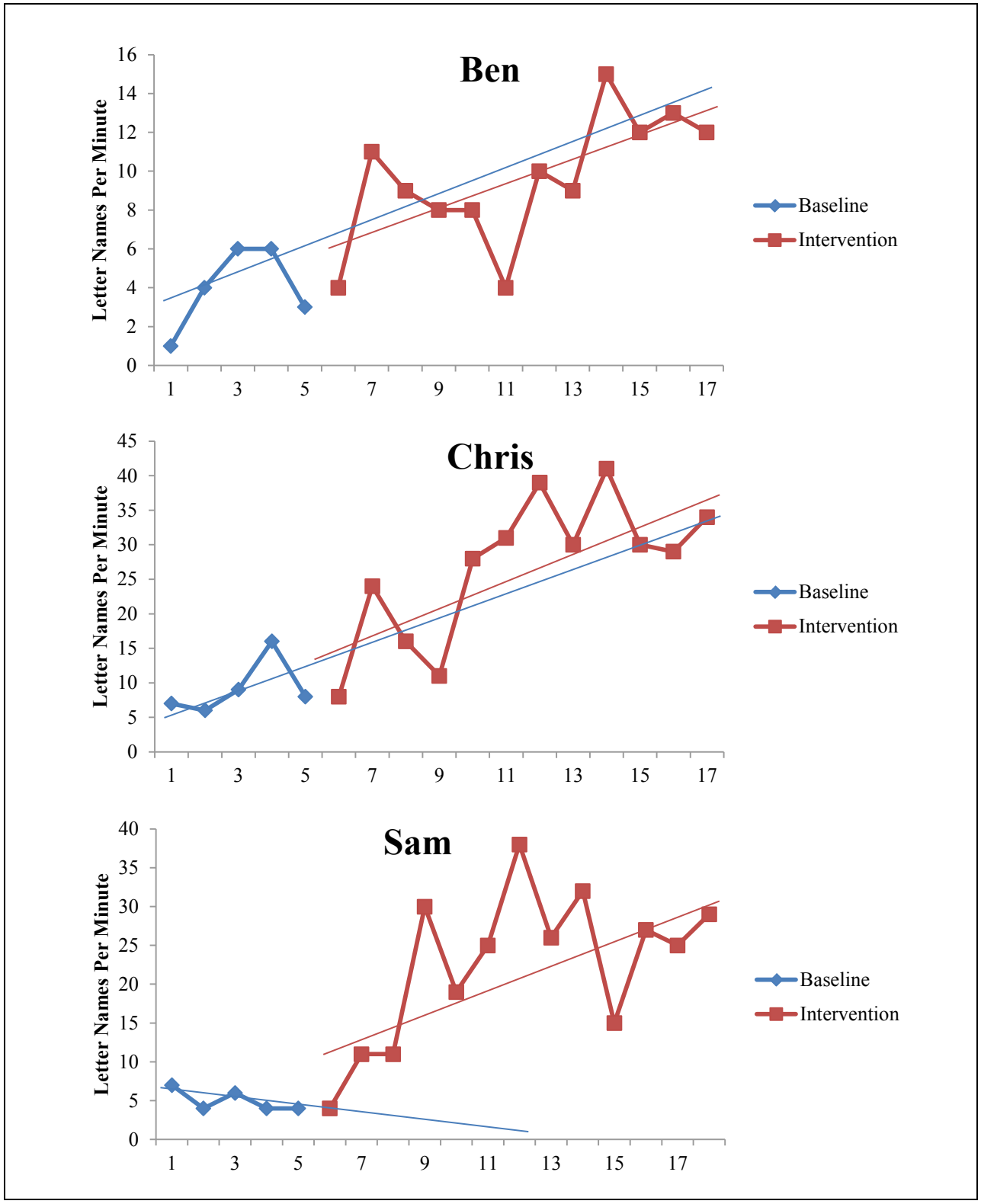


Figure 5. Group B Letter Naming Fluency with Split Middle Technique Trend Lines

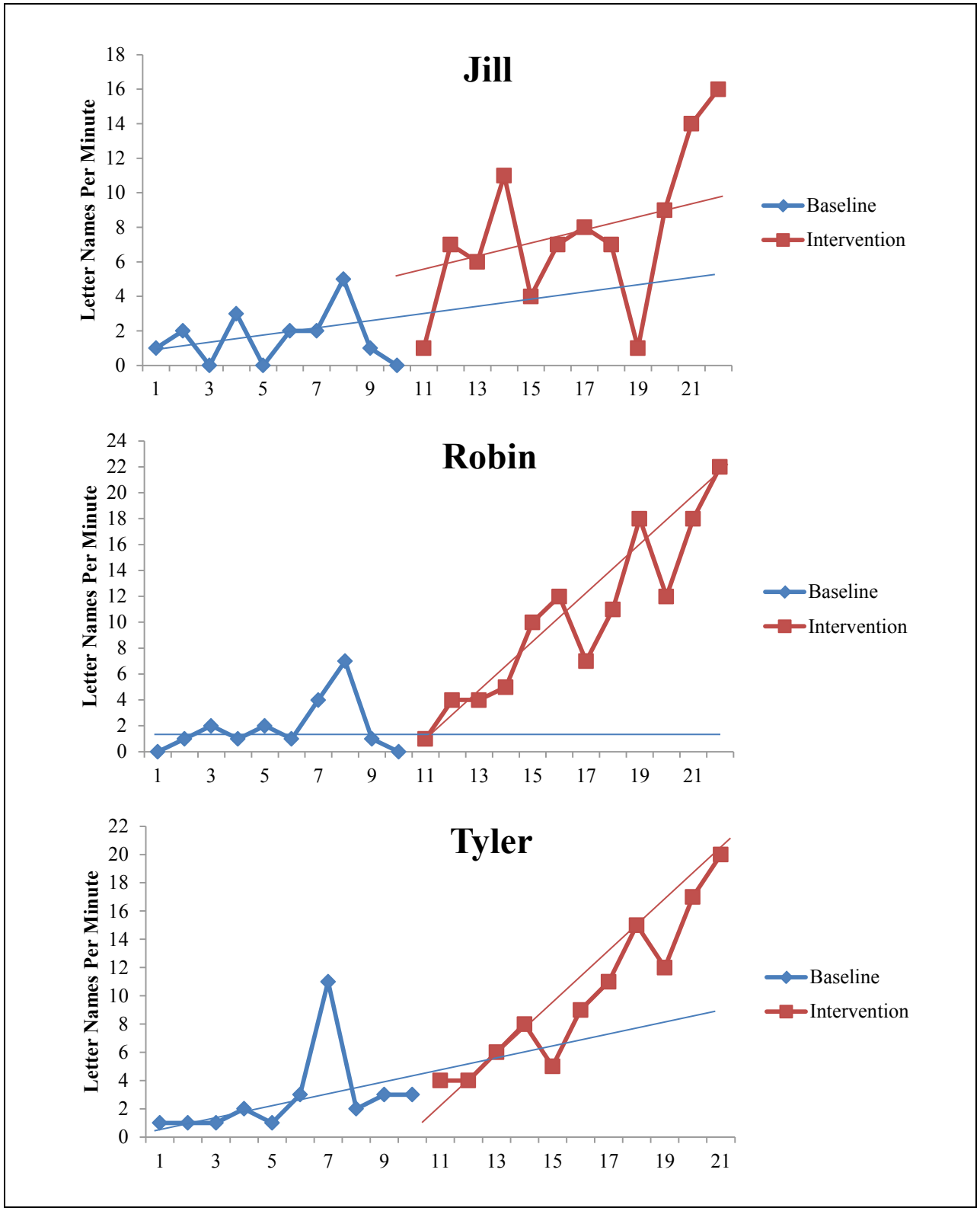
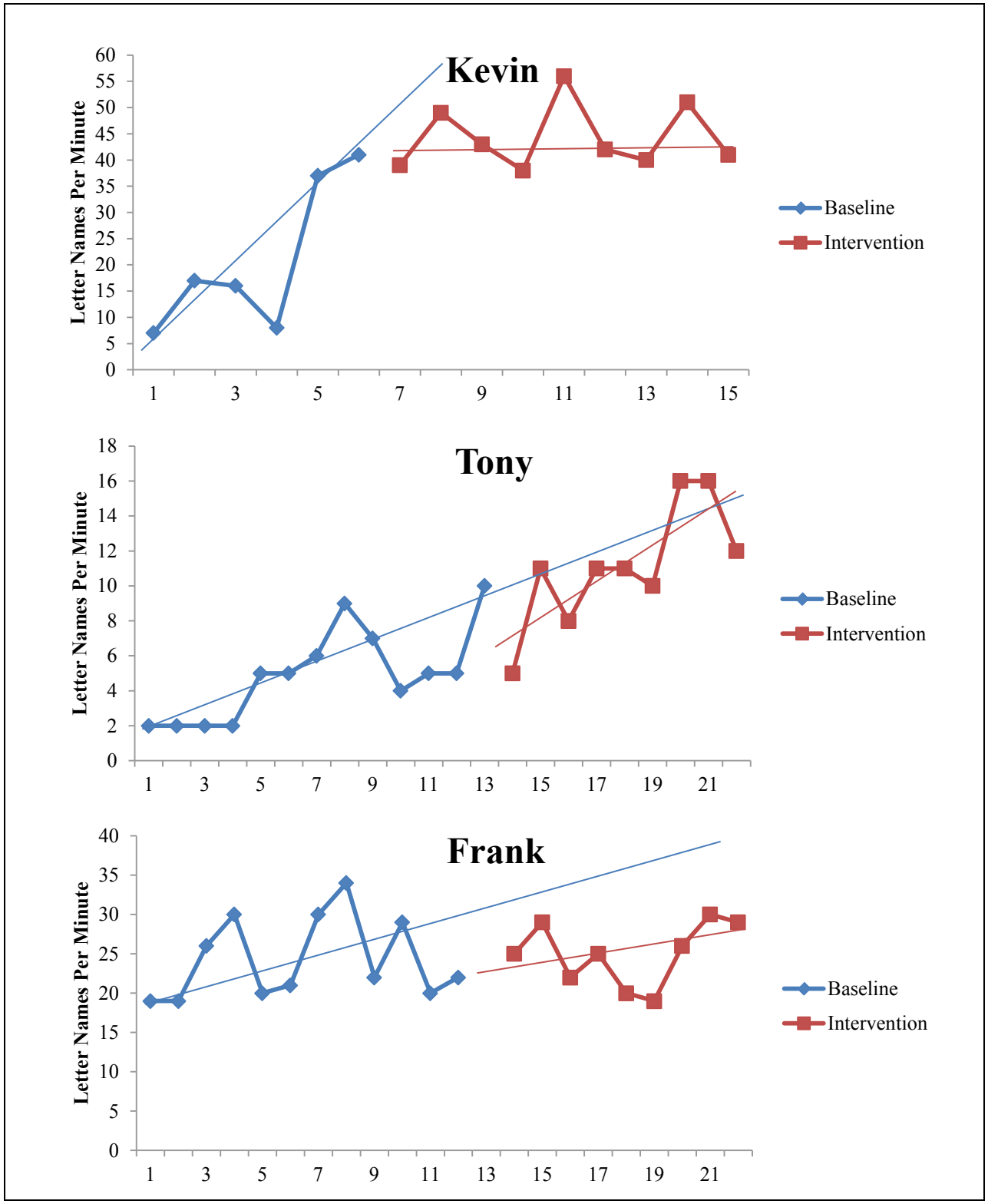


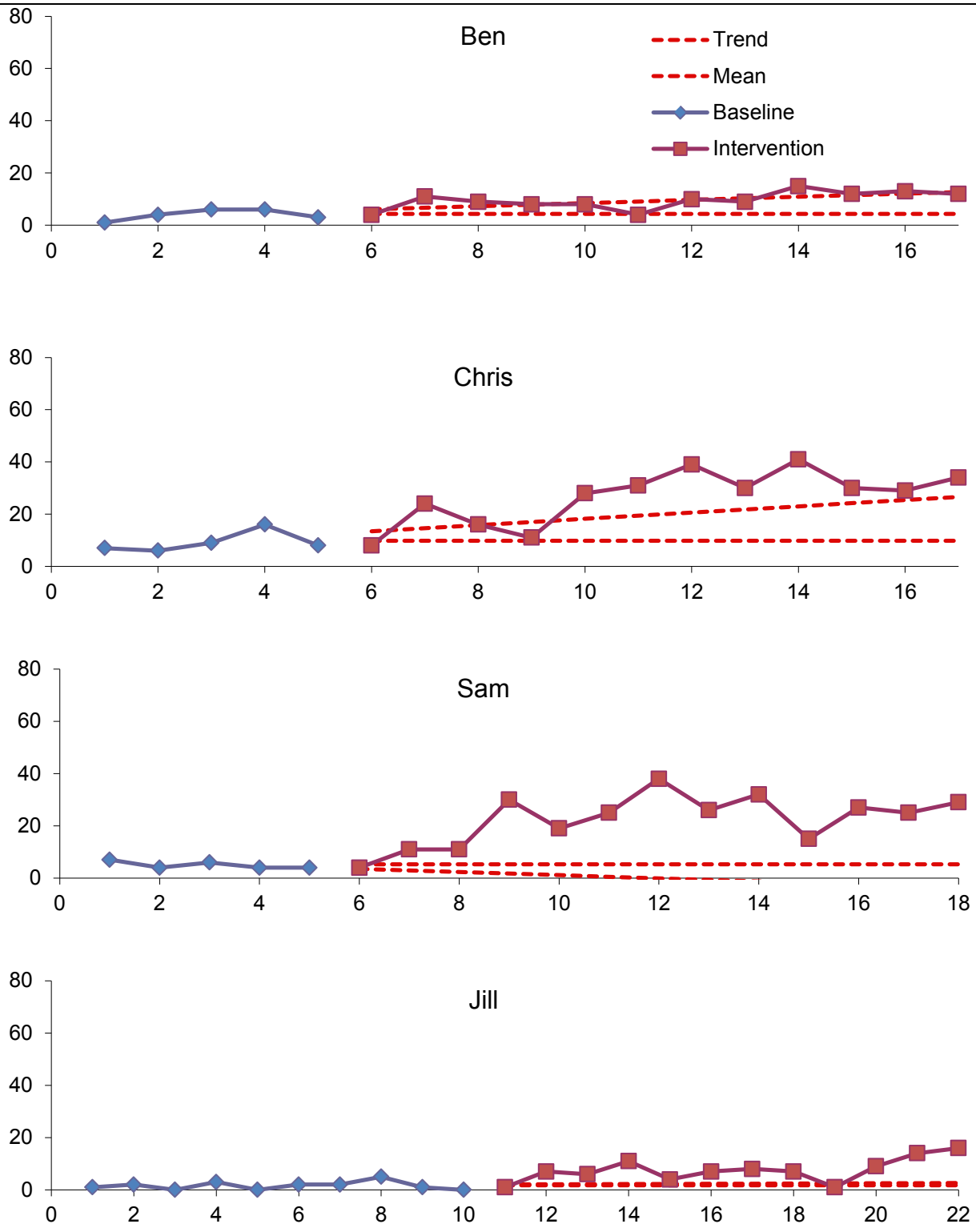
Figure 6. Group C Letter Naming Fluency with Split Middle Technique Trend Lines

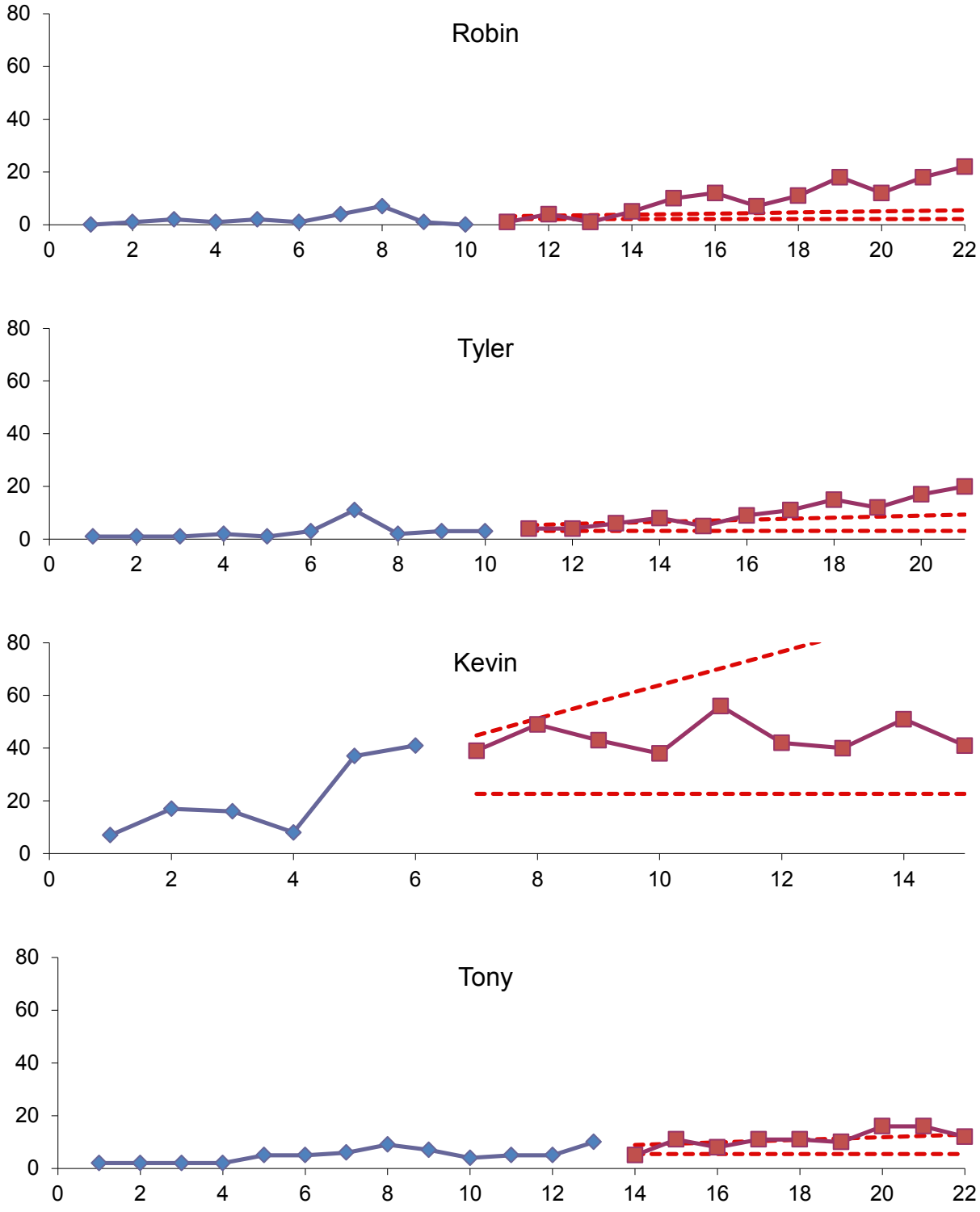


In this paragraph, data variability is reported. During baseline, data points tended to fall near the trend line for seven of the nine participants. The most variability occurred in the data of two students in Group C, Frank and Kevin. Frank's data spiked repeatedly above and below the trend line for the first six monitoring sessions. However, Frank's data showed more stability in the four monitoring sessions prior to exposure to the intervention. As previously stated, Kevin's baseline data showed a rather large spike in the data in the last two monitoring sessions of the baseline phase. The decision was made to end the baseline phase for Group C for two reasons. First, although Frank's data showed fluctuations, a consistent pattern emerged showing a relatively flat trend. Second, Frank's and Kevin's data on the other two dependent measures, LSF and FSF, had not revealed a high degree of variability. During intervention, the greatest variability occurred in Sam and Jill's data, both with repeated spikes around the trend line.

Using the conservative dual criterion (CDC) method, graphs were created for each student with the level and trend lines from the baseline phase superimposed upon the treatment phase (see Figure 7). The criteria recommended by Fisher et al. (2003) were used to determine whether or not a systematic change occurred in letter naming fluency from baseline to treatment phases. Table 3 displays the number of data points for each participant in the intervention phase, the number of data points above both the level and trend lines, and the number of data points needed to be above both lines to indicate a systematic change. The CDC method indicates that a systematic change occurred from baseline to intervention for the following students: Chris, Sam, Jill, and Robin. This corresponds to 44.4% of the participants receiving the intervention.

Figure 7. Baseline Trend and Mean Lines Superimposed upon Intervention Letter Naming Fluency Data




Robin
Tyler
Kevin
Tony

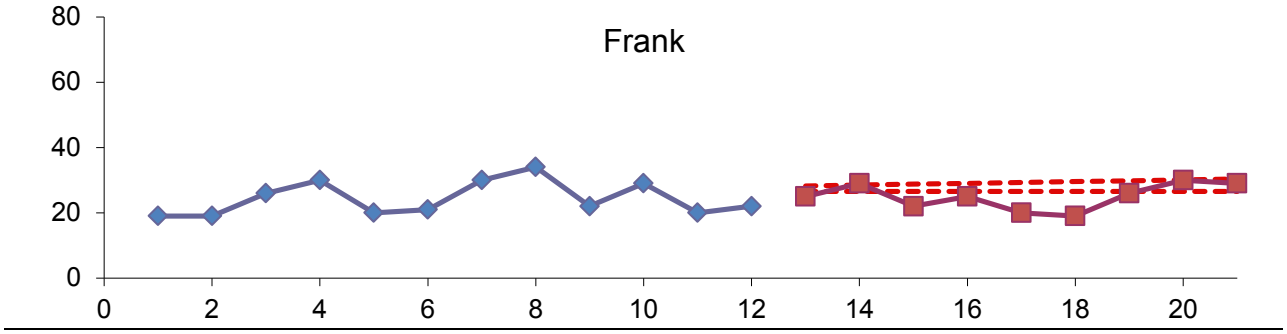


Table 3. Conservative Dual Criterion Analysis of Letter Naming Fluency

Participant	INT Points	No. Above	No. Needed
Ben	12	7	9
Chris	12	10	9
Sam	13	12	10
Jill	12	10	9
Robin	12	10	9
Tyler	11	7	9
Kevin	9	0	8
Tony	9	5	8
Frank	9	1	8

Note. INT Points = total number of intervention points in the intervention phase; No. Above = the number of intervention points above both the mean and trend lines of the baseline phase; No. Needed = the number of intervention points needed to be above both the mean and trend lines of the baseline phase to indicate that a systematic change occurred.

In this paragraph, between-phase immediacy of effect is reported for each participant. Immediacy of effect was determined by consideration of the last three data points of the baseline phase and the first three data points of the intervention phase. Using this standard, an immediacy of effect is evident for Ben, Chris, Sam, Jill, and Tyler, or 55.6% of the participants.

To analyze overlap in data, non-overlap of all pairs (NAP) was calculated for each student. The size of effects was estimated using ranges suggested by Parker & Vannest (2009) and are as follows: weak effects, 0 - .65; medium effects, .66 - .92; large effects, .93 – 1.0. Table 4 depicts the NAP for letter naming fluency for all students. Large effects occurred for Chris, Sam, Tyler, Kevin, and Tony, which represents 55.6% of participants. Medium effects occurred

for Ben, Jill, Robin, and Frank, or 44.4% of participants. Weak effects occurred for Frank, representing 11.1% of participants.

Table 4. Non-Overlap of all Pairs Effect Sizes for Letter Naming Fluency

Participant	Effect Size
Ben	0.92
Chris	0.93
Sam	0.95
Jill	0.86
Robin	0.94
Tyler	0.94
Kevin	0.94
Tony	0.93
Frank	0.54

Note. ES = effect size

Table 5 displays the consolidated results of within and between phase LNF data for each participant. In general, results were mixed regarding the effect of the intervention on students' letter name knowledge. For some students, integrating within and between phase data patterns pointed to an unequivocal result. Analysis of baseline and intervention data for four students, Chris, Sam, Jill, and Robin, indicated an increase in performance level and rate of improvement in letter name knowledge. However, a delayed effect was evident for Robin. Data for Ben, Kevin, Tony, and Frank did not indicate an increase. Integrating the different data patterns for Tyler revealed some ambiguity. Although the CDC method did not indicate a systematic change, all dimensions of within and between phase patterns did indicate a change in both level and rate of improvement. Consequently, I indicated in Table 5 that the intervention was effective for Tyler in increasing letter name knowledge. In summary, five of the nine students exhibited an increase in performance level and rate of improvement from baseline to intervention phases, which represents 55.6% of the participants.

Table 5. Integration of Within and Between Letter Naming Fluency Analyses

Name of Student	Mean Difference	Trend Difference	Variability	CDC	Immediacy of Effect	NAP	Effectiveness
Ben	5.6	Decrease	Stable	No	Yes	Medium	No
Chris	17.9	Increase	Stable	Yes	Yes	Large	Yes
Sam	17.5	Increase	Unstable in INT	Yes	Yes	Large	Yes
Jill	6.0	Increase	Unstable in INT	Yes	Yes	Medium	Yes
Robin	8.4	Increase	Stable	Yes	No	Medium	Yes
Tyler	7.3	Increase	Stable	No	Yes	Large	Yes
Kevin	23.3	Decrease	Stable	No	No	Large	No
Tony	6.2	Increase	Stable	No	No	Large	No
Frank	0.7	Decrease	Unstable in BL	No	No	Weak	No

Note. CDC = conservative dual criterion method; NAP = non-overlap of all pairs; INT = Intervention phase; BL = baseline phase.

Research Question Number Two

Is an intensive early literacy intervention effective in improving the performance level and rate of improvement of letter sound knowledge of most at-risk kindergarten students beyond baseline levels?

Table 6 displays mean level of performance as well as indicates trend level comparisons for all students in baseline and intervention phases. Mean scores are reported as letter sounds correct (LSC) per minute. All students experienced an increase in mean scores from baseline to intervention phases. The range in performance level gains varied widely across participants, from a 3.1 LSC increase for Frank to a 25.1 mean LSC increase for Kevin. Relatively small increases in performance level occurred for Ben, Jill, and Frank. More moderate increases

occurred for Robin, Tyler, and Tony, whereas Chris, Sam, and Kevin exhibited the largest increases.

Trend of the data was analyzed using the split-middle technique (see Figures 8, 9, and 10). All students' data showed a positive slope in the intervention phase, and eight of the nine students have a steeper positive slope in the intervention phase than baseline phase. Only the slope for Tony is steeper during baseline. In the trend column in Table 6, students who had more positive slopes in the intervention phase are labeled "increasing." Students who had a more positive slope during baseline are labeled "decreasing."

Table 6. Within Phase Analysis of Letter Sound Fluency Data

Participant	BL Mean	INT Mean	Trend
Ben	2.0	5.9	Increasing
Chris	0.0	14.0	Increasing
Sam	6.0	16.8	Increasing
Jill	2.5	7.3	Increasing
Robin	4.8	12.4	Increasing
Tyler	3.7	10.0	Increasing
Kevin	8.7	33.8	Increasing
Tony	3.0	8.7	Decreasing
Frank	11.7	14.8	Increasing

Note. BL = baseline; INT = intervention.

Figure 8. Group A Letter Sound Fluency with Split Middle Technique Trend Lines

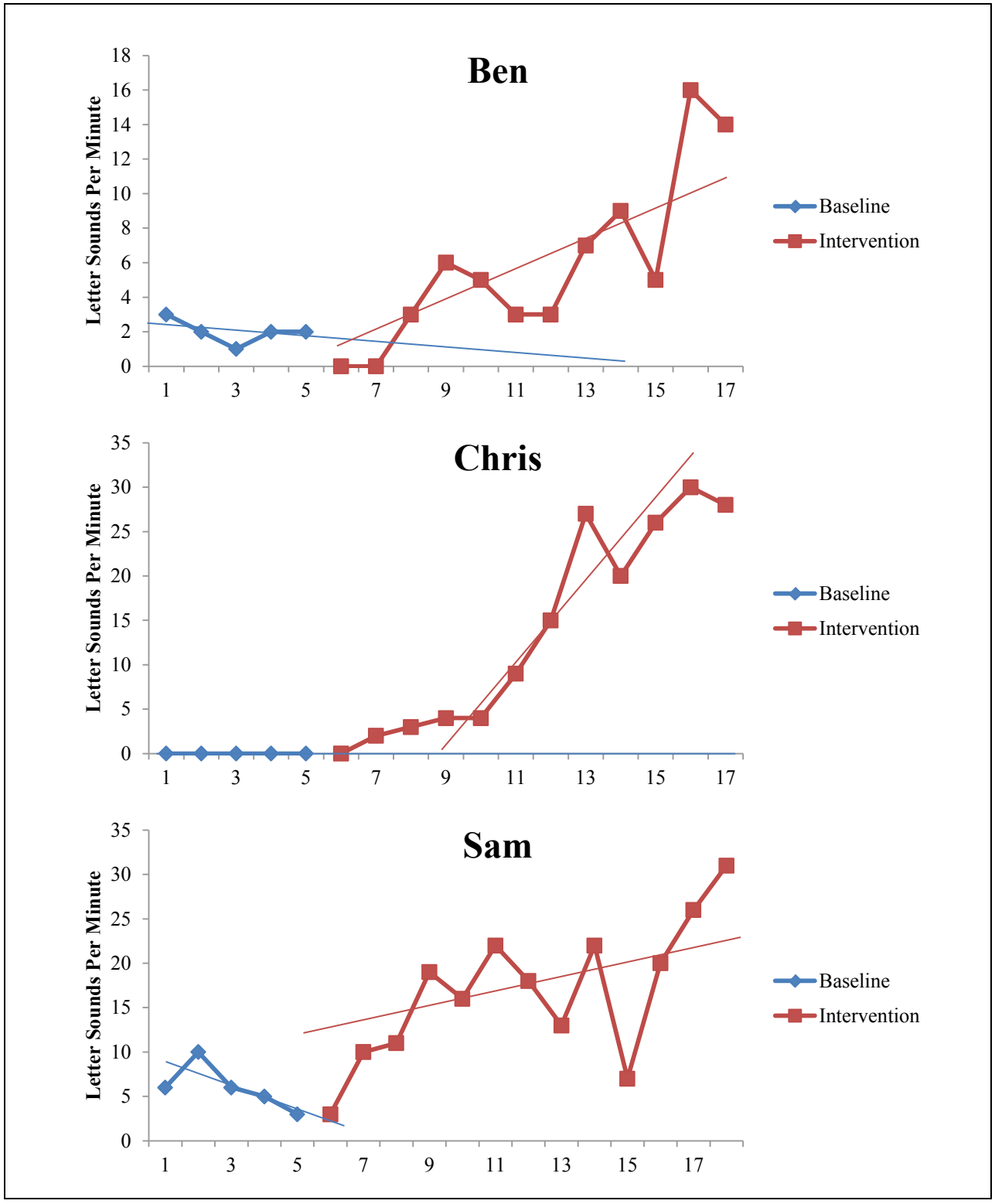


Figure 9. Group B Letter Sound Fluency with Split Middle Technique Trend Lines

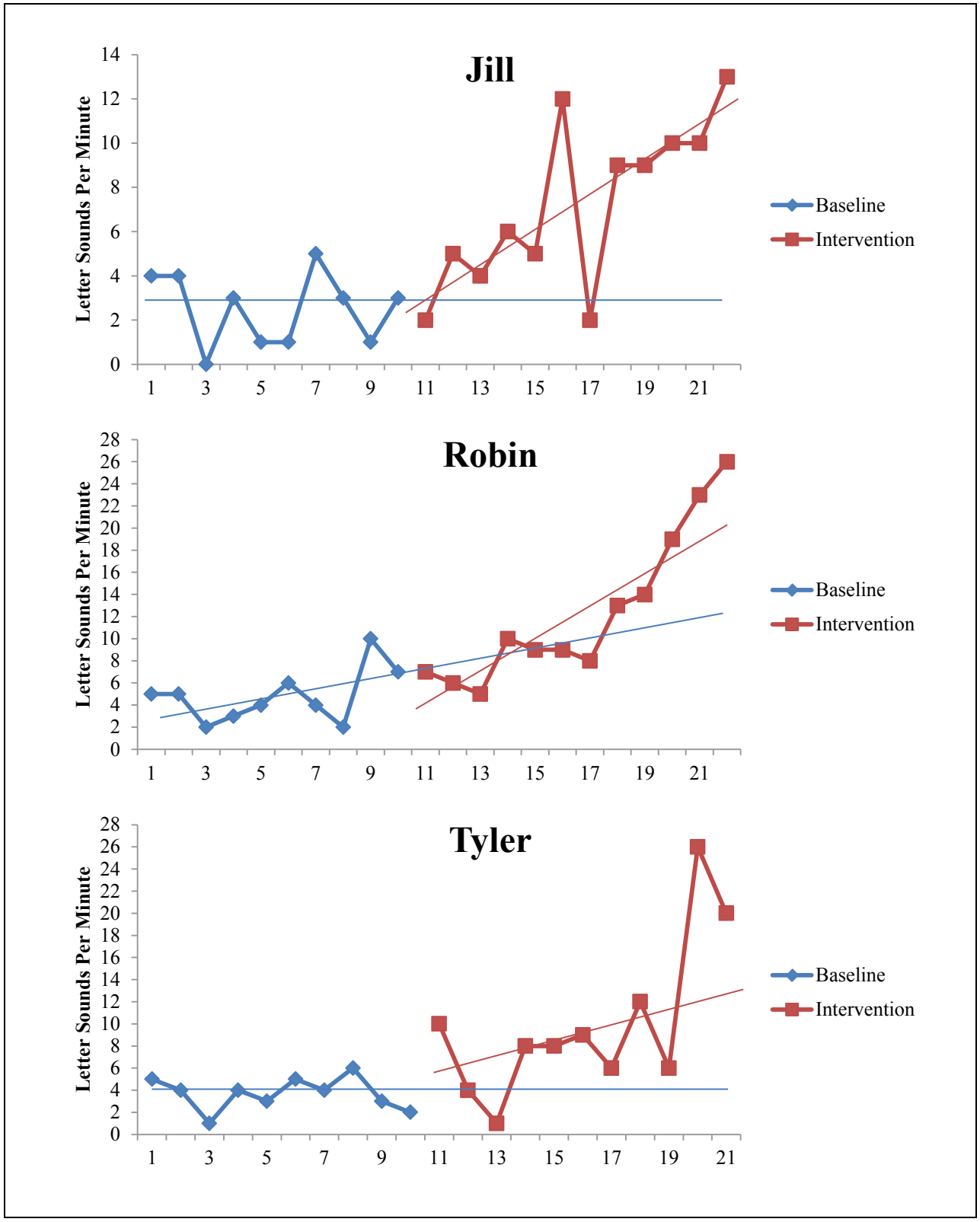
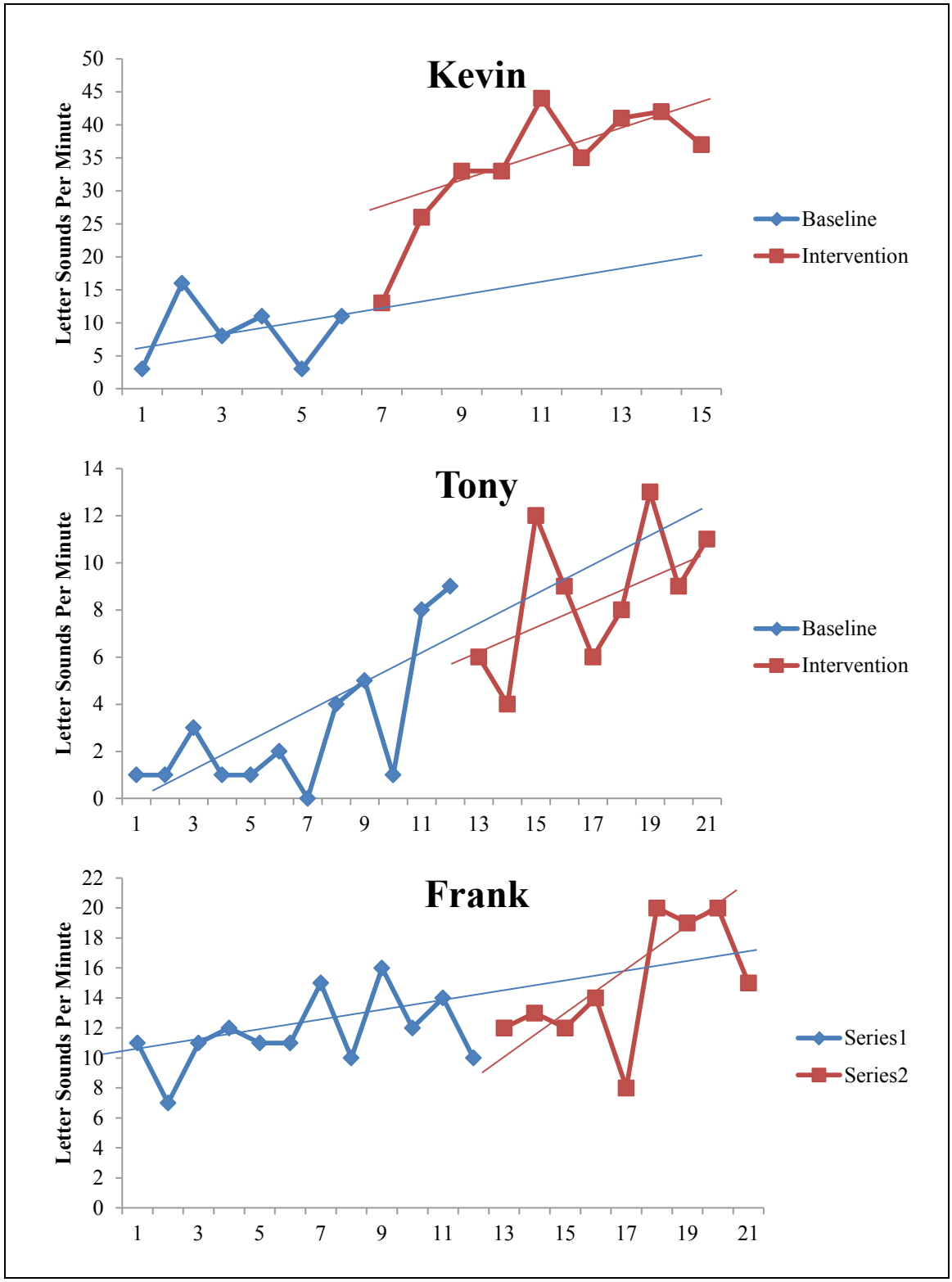


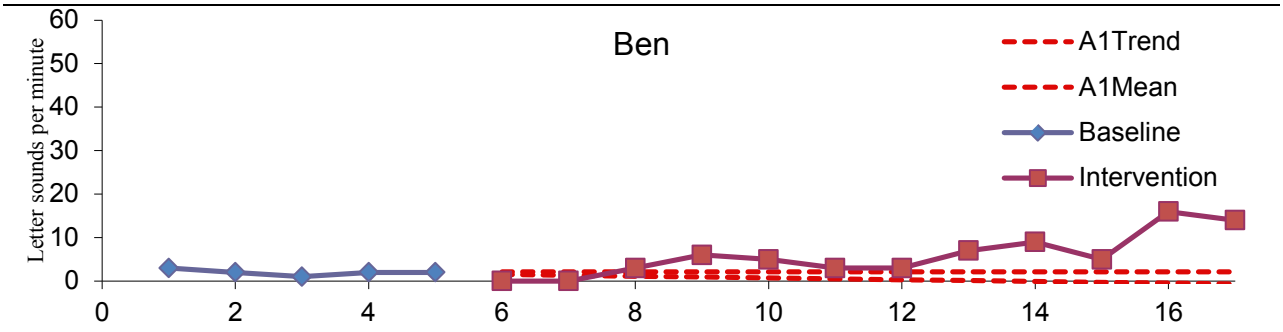
Figure 10. Group C Letter Sound Fluency with Split Middle Technique Trend Lines

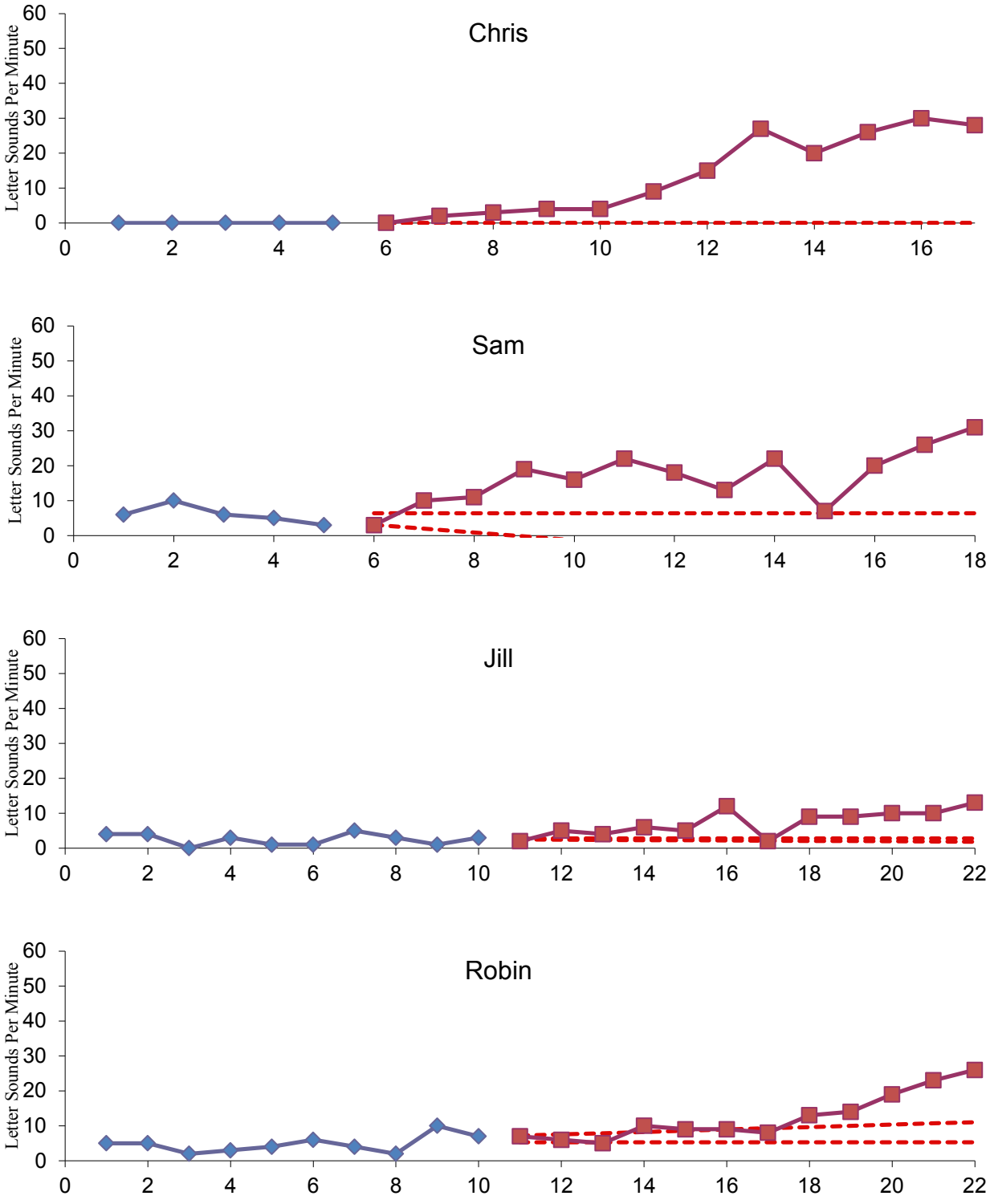


In this paragraph, data variability within phases is reported. During baseline, little variability occurred in the data for all participants. The most variability occurred in Kevin's baseline, with an initial spike in performance level followed by a return to lower levels for all other data points. In the intervention phase, data points tended to hover around the trend line for all students except Tyler, who experienced a spike in performance level at the end of intervention phase.

Using the CDC method, graphs were again created for each student with the level and trend lines from the baseline phase superimposed upon the treatment phase (see Figure 11). Table 7 displays the number of data points for each participant in the intervention phase, the number of data points above both the level and trend lines, and the number of data points needed to be above both lines to indicate a systematic change. The CDC method indicates that a systematic change occurred from baseline to intervention for the following students: Ben, Chris, Sam, Jill, Tyler, and Kevin. This corresponds to 66.7% of the participants receiving the intervention.

Figure 11. Baseline Trend and Mean Lines Superimposed upon Intervention Letter Sound Fluency Data




Chris
Letter Sounds Per Minute
60
50
40
30
20
10
0
0 2 4 6 8 10 12 14 16
Sam
Letter Sounds Per Minute
60
50
40
30
20
10
0
0 2 4 6 8 10 12 14 16 18
Jill
Letter Sounds Per Minute
60
50
40
30
20
10
0
0 2 4 6 8 10 12 14 16 18 20 22
Robin
Letter Sounds Per Minute
60
50
40
30
20
10
0
0 2 4 6 8 10 12 14 16 18 20 22

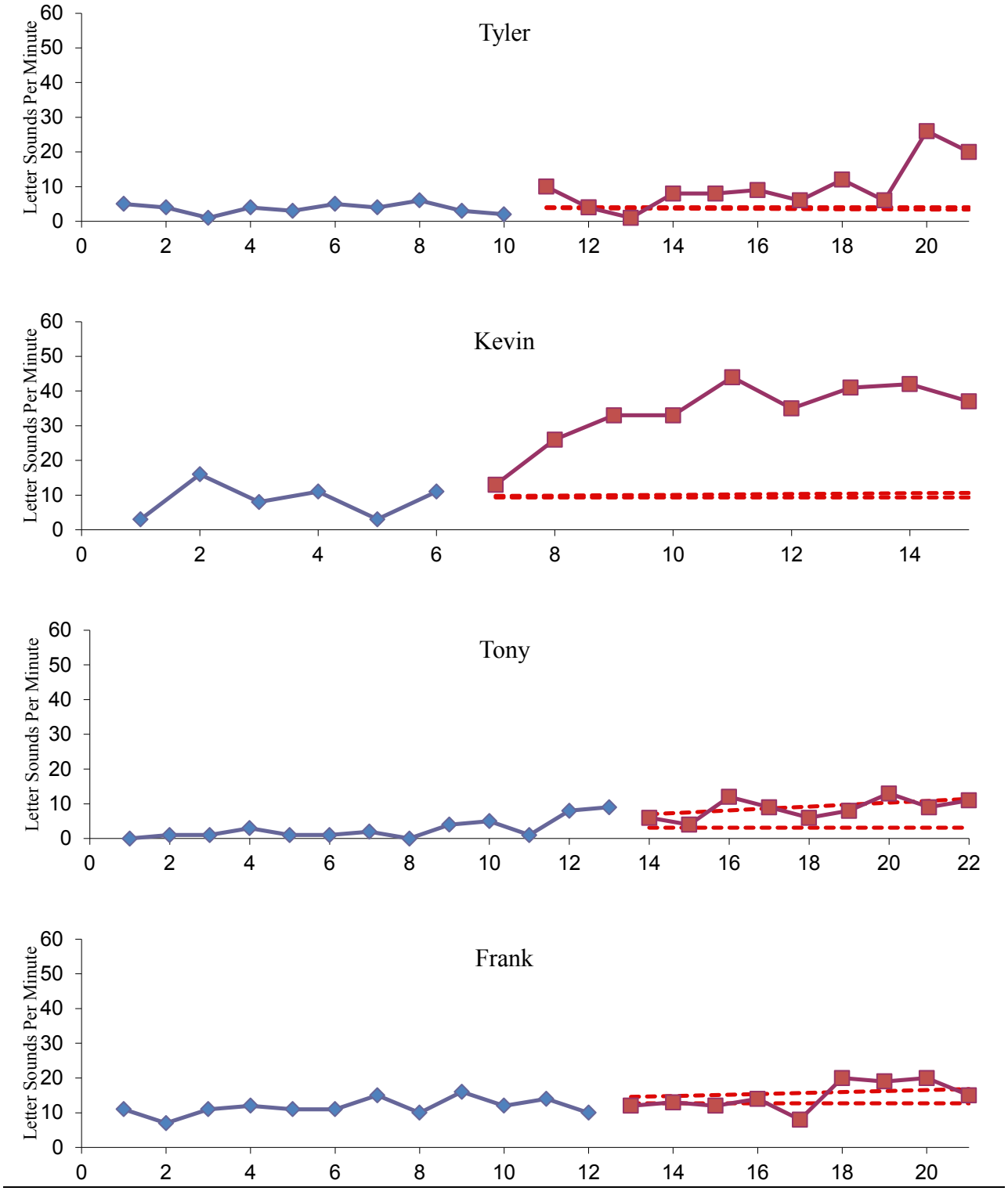

Tyler
Letter Sounds Per Minute
0 10 20 30 40 50 60
0 2 4 6 8 10 12 14 16 18 20
Kevin
Letter Sounds Per Minute
0 10 20 30 40 50 60
0 2 4 6 8 10 12 14
Tony
Letter Sounds Per Minute
0 10 20 30 40 50 60
0 2 4 6 8 10 12 14 16 18 20 22
Frank
Letter Sounds Per Minute
0 10 20 30 40 50 60
0 2 4 6 8 10 12 14 16 18 20

Table 7. Conservative Dual Criterion Analysis of Letter Sound Fluency

Participant	No. of INT Points	No. Above	No. Needed
Ben	12	10	9
Chris	12	11	9
Sam	13	12	10
Jill	12	10	9
Robin	12	8	9
Tyler	11	9	9
Kevin	9	9	8
Tony	9	3	8
Frank	9	3	8

Note. INT Points = total number of intervention points in the intervention phase; No. Above = the number of intervention points above both the mean and trend lines of the baseline phase; No. Needed = the number of intervention points needed to be above both the mean and trend lines of the baseline phase to indicate that a systematic change occurred.

In this paragraph, between-phase immediacy of effect is reported for each participant. Immediacy of effect was determined by consideration of the last three data points of the baseline phase and the first three data points of the intervention phase. Using this standard, an immediacy of effect was evident for Chris, Sam, Jill, and Kevin, or 44.4% of the participants.

To analyze overlap in data, non-overlap of all pairs was calculated for each student. The size of effects was estimated using ranges suggested by Parker & Vannest (2009) and are as follows: weak effects, 0 - .65; medium effects, .66 - .92; large effects, .93 – 1.0. Table 8 depicts the NAP for letter sound fluency for all students. Large effects occurred for Chris and Kevin, or 22.2% of participants. Medium effects occurred for Ben, Sam, Jill, Robin, Tyler, Tony, and Frank, or 77.8% of participants.

Table 8. Non-Overlap of all Pairs Effect Sizes for Letter Sound Fluency

Participant	Effect Size
Ben	0.80
Chris	0.96
Sam	0.91
Jill	0.86
Robin	0.90
Tyler	0.86
Kevin	0.98
Tony	0.91
Frank	0.76

Note. ES = effect size

Table 9 displays the consolidated results of within and between phase LSF data for each participant. Similar to letter name knowledge, results were mixed regarding the effect of the intervention on students' letter sound knowledge. For some students, integrating within and between phase data patterns again pointed to an unequivocal result. Analysis of baseline and intervention data for four students, Chris, Sam, Jill, and Kevin, indicated an increase in performance level and rate of improvement in letter sound knowledge. Data for Tony and Frank did not indicate an increase. Integrating the different data patterns for Ben, Robin, and Tyler revealed some ambiguity.

For Ben, most data patterns indicated an effect of the intervention. For example, the split middle technique indicated a clear, positive increase in rate of improvement, little variability existed in his scores, the CDC method indicated a systematic change, and a medium effect size was present according to NAP. On the contrary, Ben's performance level increase was relatively small and immediacy of effect using the last three data points was not evident. However, data points from the first three weeks of intervention did indicate a clear change in level and trend. Consequently, I considered the intervention effective for increasing Ben's letter sound knowledge.

For Robin, the following dimensions indicate an effect: a moderate difference in mean performance level, a clear and positive difference in rate of improvement, little variability in the data, and a medium effect size of NAP. Using the CDC method, Robin was one data point shy of the minimum criteria indicating a systematic change. Additionally, a clear change in level and rate of improvement is not evident until week six of the intervention phase, leaving ambiguity in terms of the cause of this difference. Consequently, I indicated in Table 9 that the intervention was not effective for Robin in increasing letter sound knowledge.

For Tyler, the following dimensions indicate an effect: a moderate difference in mean performance level, a clear and positive difference in rate of improvement, a systematic change according to the CDC method, and a medium effect size of NAP. Conflicting data was present in terms of some instability of data during intervention phase and no apparent immediacy of effect using the three-data-point standard. The large spike at the end of the intervention phase could perhaps be explained with a decrease in Tyler's behavioral difficulties. During baseline and intervention phases, Tyler had 31 disciplinary referrals. Only two of these occurred during the last four weeks of the intervention phase. Similar to Ben, data points from the first three weeks of intervention did indicate a clear change in level and trend. Consequently, I considered the intervention effective for increasing Tyler's letter sound knowledge. In summary, six of the nine students exhibited an increase in performance level and trend from baseline to intervention phases, which represents 66.7% of the participants.

Table 9. Integration of Within and Between Phase Letter Sound Fluency Analyses

Name of Student	Mean Difference	Trend Difference	Variability	CDC	Immediacy of Effect	NAP	Effectiveness
Ben	3.9	Increase	Stable	Yes	No	Medium	Yes
Chris	14.0	Increase	Stable	Yes	Yes	Large	Yes
Sam	10.8	Increase	Stable	Yes	Yes	Medium	Yes
Jill	4.8	Increase	Stable	Yes	Yes	Medium	Yes
Robin	7.6	Increase	Stable	No	No	Medium	No
Tyler	6.3	Increase	Unstable in INT	Yes	No	Medium	Yes
Kevin	25.1	Increase	Stable	Yes	Yes	Large	Yes
Tony	5.7	Decrease	Stable	No	No	Medium	No
Frank	3.1	Increase	Stable	No	No	Medium	No

Note. CDC = conservative dual criterion method; NAP = non-overlap of all pairs; INT = Intervention phase.

Research Question Number Three

Is an intensive early literacy intervention effective in improving the performance level and rate of improvement of phonemic awareness, as measured by first sound identification, of most at-risk kindergarten students beyond baseline levels?

Table 10 displays mean level of performance as well as indicates trend level comparisons for all students in baseline and intervention phases. Mean scores are reported as first sounds correct (FSC) per minute. All students exhibited an increase in mean scores from baseline to intervention phases. The range in performance level gains varied from a 10.2 mean FSC increase for Robin to a 39.9 mean FSC increase for Kevin. Smaller increases in performance level occurred for Sam, Robin, Tyler, and Tony. More moderate increases occurred for Ben, Chris, Jill, and Frank.

Trend of the data was analyzed using the split-middle technique (see Figures 12, 13, and 14). Seven students had a positive slope in the intervention phase. Data for Tony and Frank showed a negative slope during intervention. Six of the nine students have a steeper positive

slope in the intervention phase than baseline phase. The slopes for Robin, Tony, and Frank are all steeper during baseline. In the trend column in Table 10, students who had more positive slopes in the intervention phase are labeled "increasing." Students who had a more positive slope during baseline are labeled "decreasing."

Table 10. Within Phase Analysis of First Sound Fluency Data

Participant	BL Mean	INT Mean	Trend
Ben	3.8	26.8	Increasing
Chris	1.4	21.6	Increasing
Sam	0.0	15.4	Increasing
Jill	2.2	24.3	Increasing
Robin	19.6	29.8	Decreasing
Tyler	29.7	42.7	Increasing
Kevin	3.3	43.2	Increasing
Tony	31.3	48.2	Decreasing
Frank	17.6	38.0	Decreasing

Note. BL = baseline; INT = intervention.

Figure 12. Group A First Sound Fluency with Split Middle Technique Trend Lines

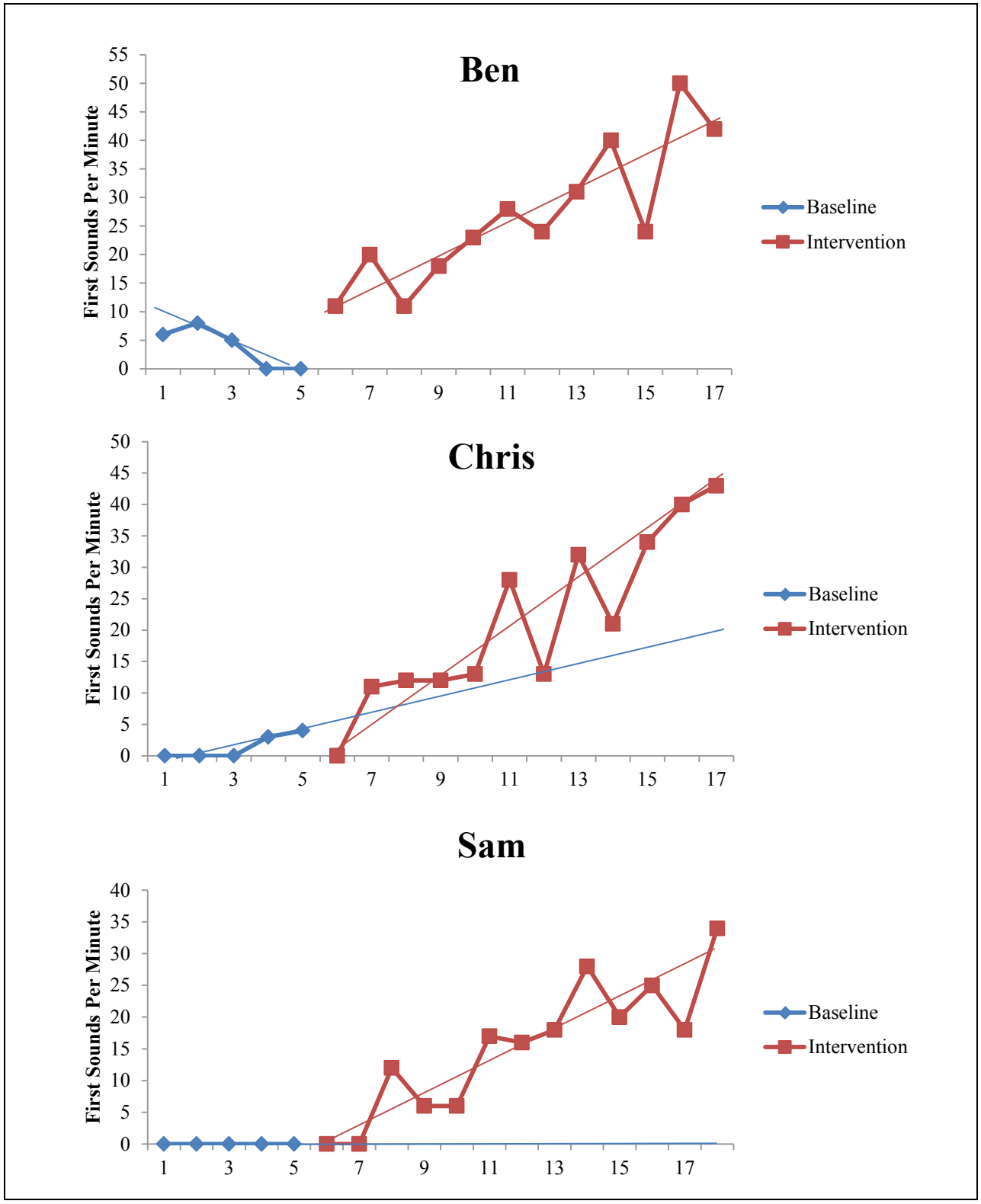


Figure 13. Group B First Sound Fluency with Split Middle Technique Trend Lines

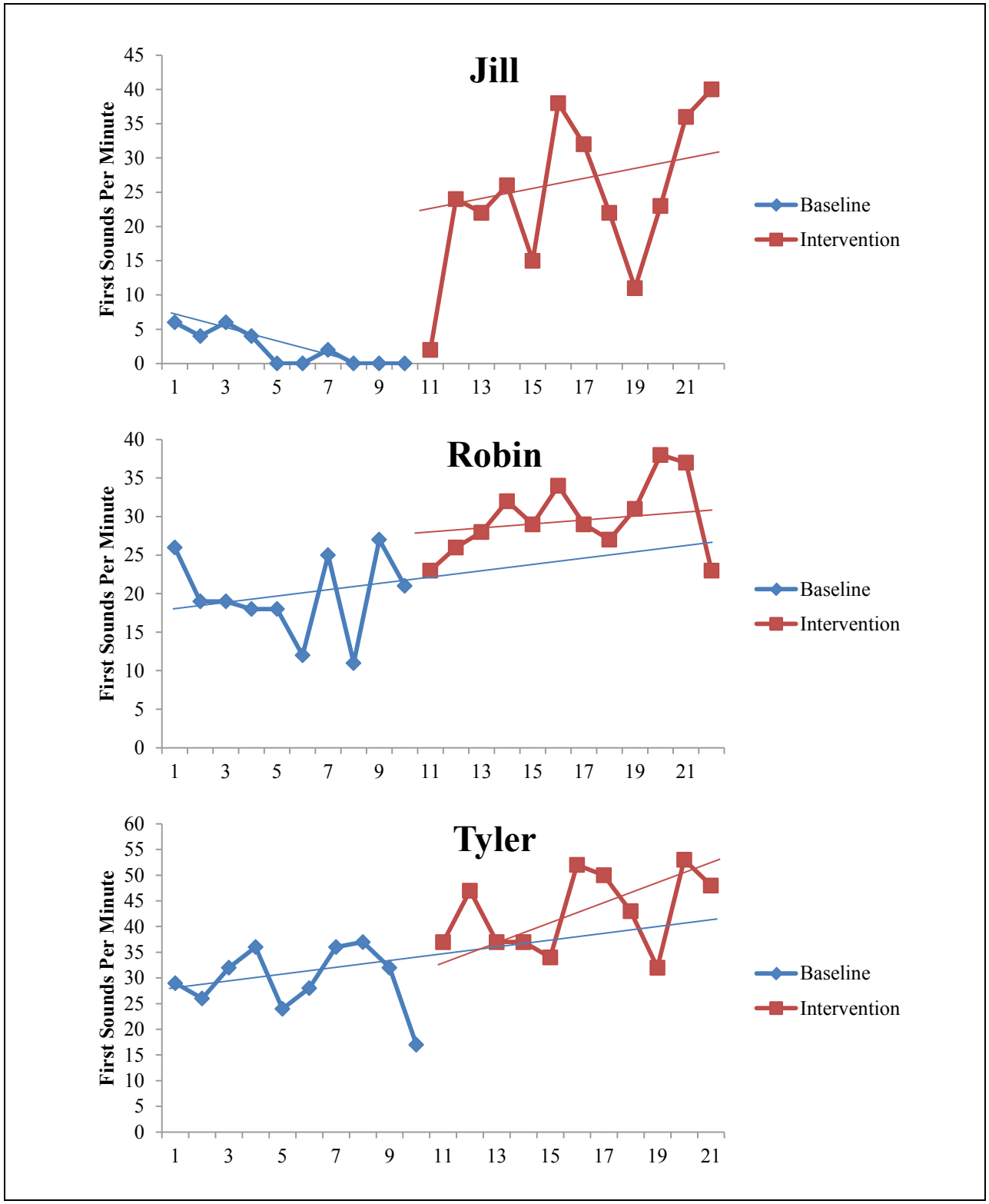
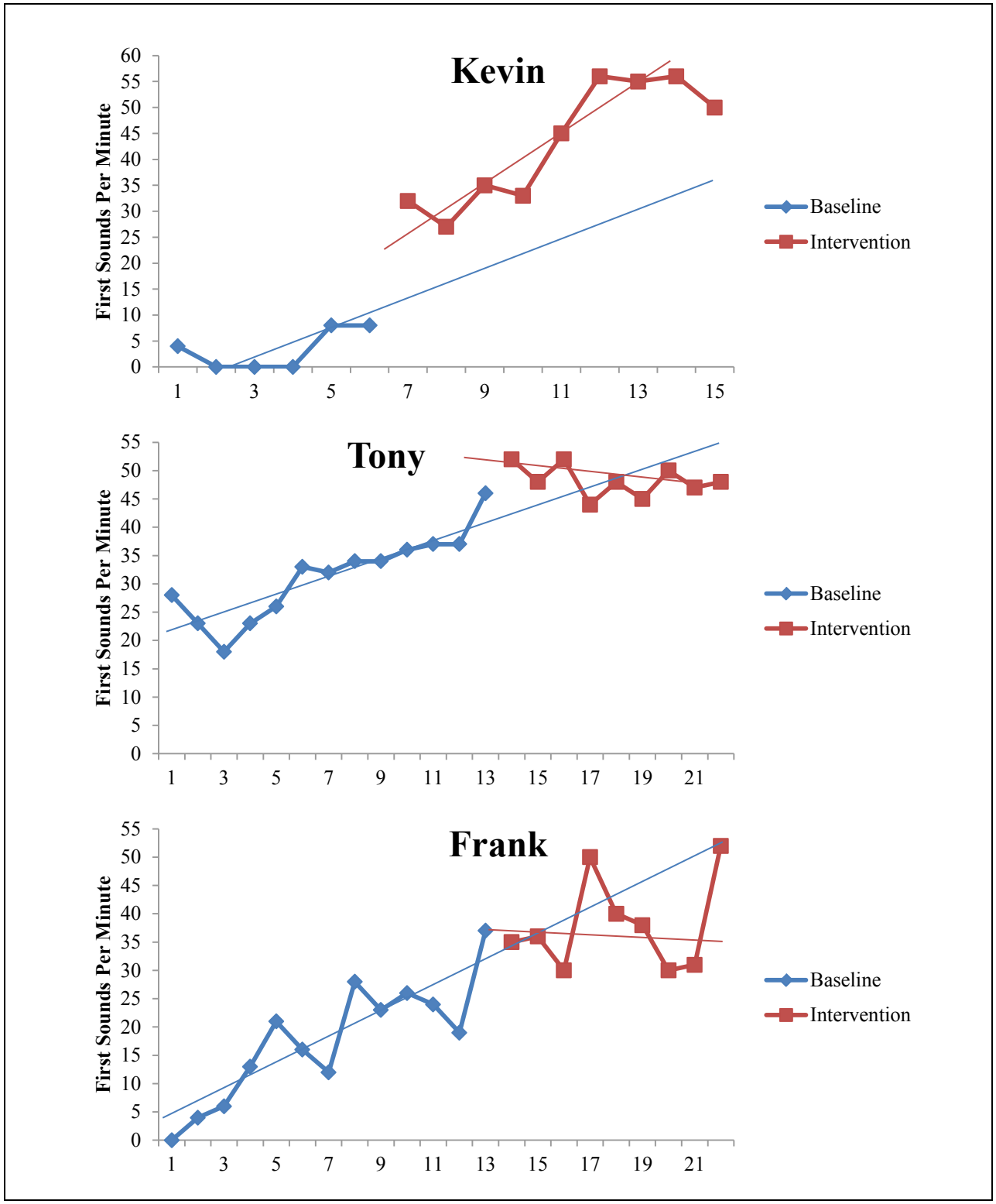


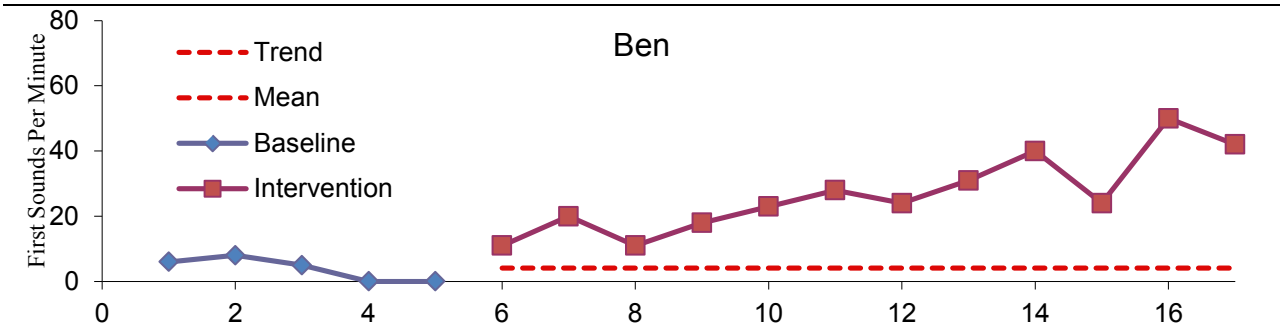
Figure 14. Group C First Sound Fluency with Split Middle Technique Trend Lines

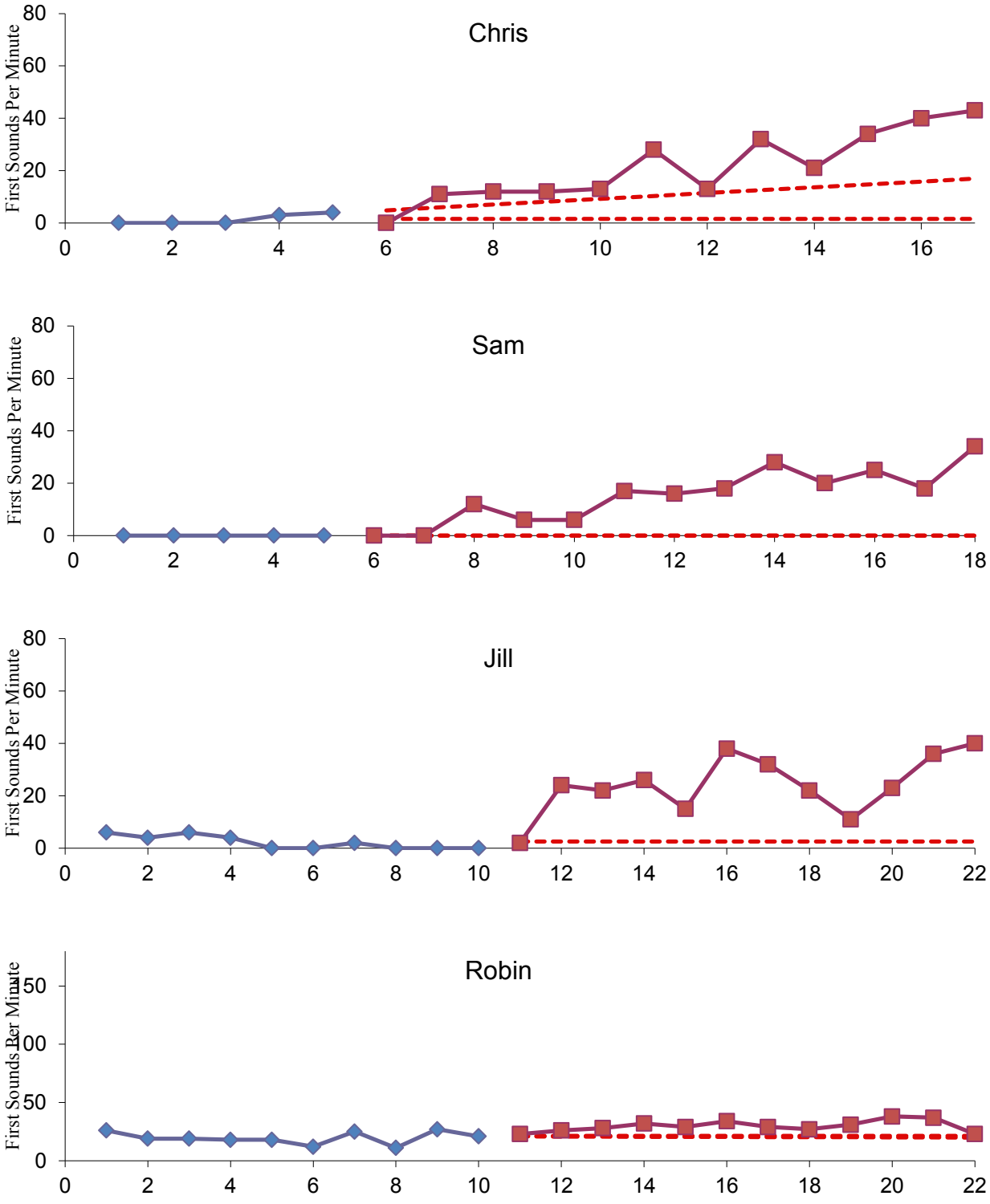


In this paragraph, data variability within phases is reported. During baseline, variability was present in the data for Robin and Tyler, showing repeated up and down spikes. Additionally, Tony and Frank's data showed obvious upward trends. In the intervention phase, variability is seen for Jill, Tyler, and Frank. The data for all three students showed repeated up and down spikes in performance level, although to a lesser extent for Frank.

Using the CDC method, graphs were created for each student with the level and trend lines from the baseline phase superimposed upon the treatment phase (see Figure 15). Table 11 displays the number of data points for each participant in the intervention phase, the number of data points above both the level and trend lines, and the number of data points needed to be above both lines to indicate a systematic change. The CDC method indicates that a systematic change occurred from baseline to intervention for the following students: Ben, Chris, Sam, Jill, Robin, Tyler, and Kevin. This corresponds to 77.8% of the participants receiving the intervention.

Figure 15. Baseline Trend and Mean Lines Superimposed upon Intervention First Sound Fluency Data




Chris
First Sounds Per Minute
0
20
40
60
80
0
2
4
6
8
10
12
14
16
Sam
First Sounds Per Minute
0
20
40
60
80
0
2
4
6
8
10
12
14
16
18
Jill
First Sounds Per Minute
0
20
40
60
80
0
2
4
6
8
10
12
14
16
18
20
22
Robin
First Sounds Per Minute
0
50
100
150
0
2
4
6
8
10
12
14
16
18
20
22

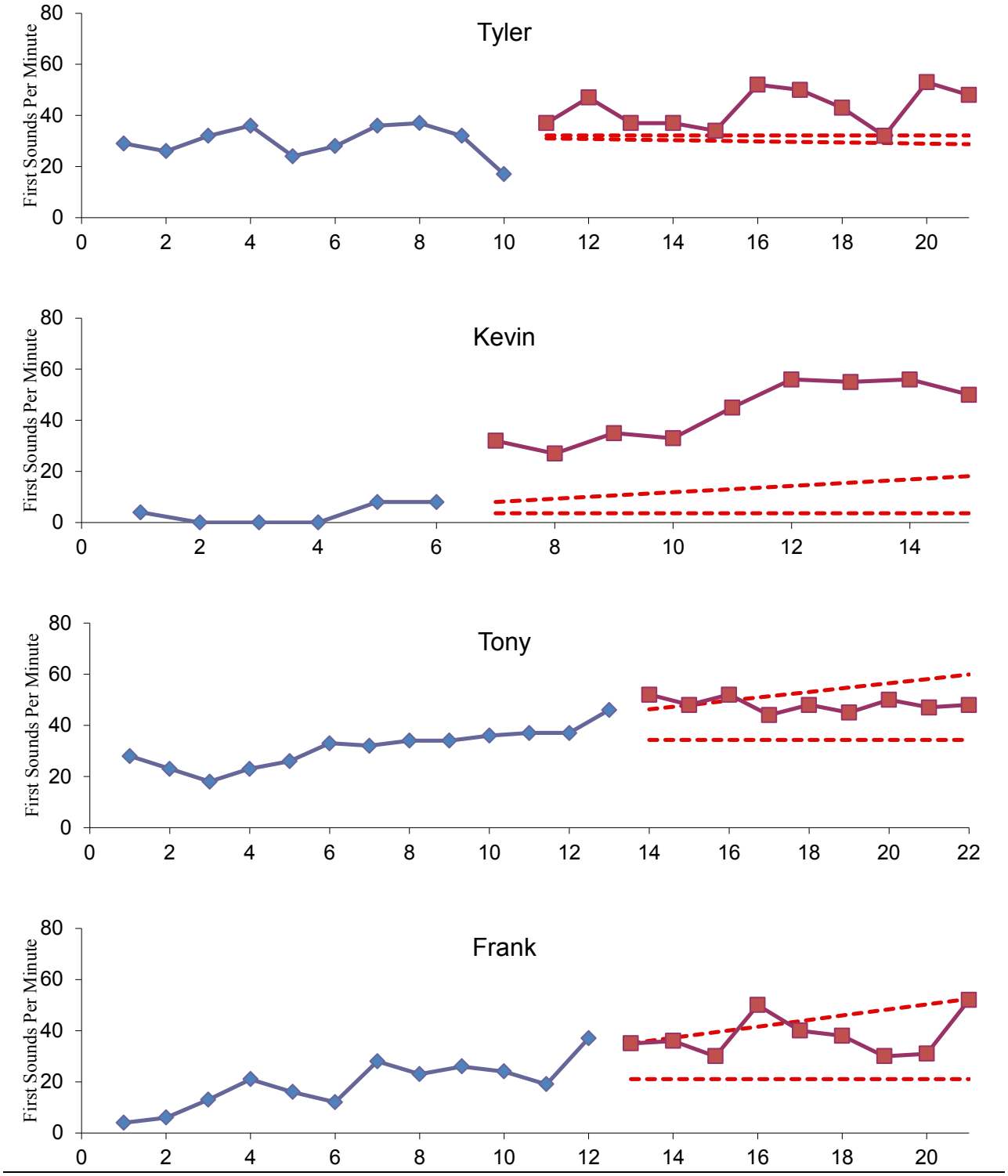

Tyler
First Sounds Per Minute
0
20
40
60
80
0 2 4 6 8 10 12 14 16 18 20
Kevin
First Sounds Per Minute
0
20
40
60
80
0 2 4 6 8 10 12 14
Tony
First Sounds Per Minute
0
20
40
60
80
0 2 4 6 8 10 12 14 16 18 20 22
Frank
First Sounds Per Minute
0
20
40
60
80
0 2 4 6 8 10 12 14 16 18 20

Table 11. Conservative Dual Criterion Analysis of First Sound Fluency

Participant	No. of INT Points	No. Above	No. Needed
Ben	12	12	9
Chris	12	11	9
Sam	13	11	10
Jill	12	11	9
Robin	12	12	9
Tyler	11	10	9
Kevin	9	9	8
Tony	9	3	8
Frank	9	1	8

Note. INT Points = total number of intervention points in the intervention phase; No. Above = the number of intervention points above both the mean and trend lines of the baseline phase; No. Needed = the number of intervention points needed to be above both the mean and trend lines of the baseline phase to indicate that a systematic change occurred.

In this paragraph, between-phase immediacy of effect is reported for each participant. Immediacy of effect was determined by consideration of the last three data points of the baseline phase and the first three data points of the intervention phase. Using this standard, an immediacy of effect was evident for Ben, Chris, Sam, Jill, Kevin, and Tony, or 66.7% of the participants. However, Tony's last data point in baseline spiked upward in performance level, calling into question whether the independent variable was the cause of this effect. No immediacy of effect was present for Robin, Tyler, or Frank.

To analyze overlap in data, non-overlap of all pairs was calculated for each student. The size of effects was estimated using ranges suggested by Parker & Vannest (2009) and are as follows: weak effects, 0 - .65; medium effects, .66 - .92; large effects, .93 – 1.0. Table 12 depicts the NAP for first sound fluency for all students. Large effects were seen for Ben, Chris, Sam, Jill, Robin, Kevin, Tony, and Frank, or 88.9% of participants. A medium effect was seen for Tyler, or 11.1% of participants.

Table 12. Non-Overlap of all Pairs Effect Sizes for First Sound Fluency

Participant	Effect Size
Ben	1.00
Chris	0.94
Sam	0.93
Jill	0.96
Robin	0.95
Tyler	0.92
Kevin	1.00
Tony	0.98
Frank	0.95

Note. ES = effect size

Table 13 displays the consolidated results of within and between phase FSF data for each participant. Similar to letter name and letter sound knowledge, results were mixed regarding the effect of the intervention on students' first sound identification skills. For some students, integrating within and between phase data patterns again pointed to an unequivocal result. Analysis of baseline and intervention data for five students, Ben, Chris, Sam, Jill, and Kevin, indicated an increase in performance level and rate of improvement in first sound identification. Data for Tony and Frank did not indicate an increase in rate of improvement. Integrating the different data patterns for Robin and Tyler revealed some ambiguity.

For Robin, the following dimensions indicated an effect: a clear change in performance level, a systematic change according to the CDC method, and a large effect size using the formula for NAP. On the contrary, trend lines created using the split middle technique showed a steeper positive trend during baseline. Additionally, data points showed some variability during baseline and no immediacy of effect was apparent. Although the split middle technique produced a steeper trend in baseline than intervention, the CDC method showed all data points (i.e. 12 of 12) in the intervention phase above the mean and trend lines of the baseline phase.

Additionally, a change in performance level and trend was apparent after three weeks of the intervention. Consequently, I considered the intervention effective for Robin.

For Tyler, the following dimensions indicated an effect: a change in performance level, a steeper trend in the intervention phase, a systematic change according to the CDC method, and a medium effect size using the formula for NAP. However, Tyler's data did show variability in both phases and no immediacy of effect. An effect did not become apparent until after three weeks of intervention. After this point, wide fluctuations in the data were also present. Both of these patterns created ambiguity in terms of the cause of performance level and trend differences across phases. Despite positive indications of an effect from the split middle technique and the CDC method, I considered the intervention ineffective for Tyler. In summary, six of the nine students exhibited differences in performance level and rate of improvement from baseline to intervention phases.

Table 13. Integration of Within and Between First Sound Fluency Analyses

Name of Student	Mean Difference	Trend Difference	Variability	CDC	Immediacy of Effect	NAP	Effectiveness
Ben	23.0	Increase	Stable	Yes	Yes	Large	Yes
Chris	20.2	Increase	Stable	Yes	Yes	Large	Yes
Sam	15.4	Increase	Stable	Yes	Yes	Large	Yes
Jill	22.1	Increase	Unstable in INT	Yes	Yes	Large	Yes
Robin	10.2	Decrease	Unstable in BL	Yes	No	Large	Yes
Tyler	13.0	Increase	Unstable in BL and INT	Yes	No	Medium	No
Kevin	39.9	Increase	Stable	Yes	Yes	Large	Yes
Tony	16.9	Decrease	Stable	No	Yes	Large	No
Frank	20.4	Decrease	Unstable in INT	No	No	Large	No

Note. CDC = conservative dual criterion method; NAP = non-overlap of all pairs; INT = Intervention phase; BL = baseline phase.

Follow-up Data

Students in Groups A and B were administered the three dependent measures three and four weeks after intervention. Follow-up data was not collected for Group C because of time constraints. Data was collected to observe the degree to which students were maintaining their alphabet knowledge and phonological awareness skill levels post-intervention phase. With only two data points, data was not intended to indicate another phase. In addition, data was not used in analyses to determine a cause and effect relationship between the independent and dependent variables or intervention effectiveness.

To be systematic, I compared follow-up data to the last three data points of the intervention phase. For a visual display of follow-up data, see Figures 1, 2, and 3 at the beginning of this chapter. Ben's follow-up data indicated similar levels of performance for both LNF and FSF and a substantial drop in LSF as compared to end-of-intervention data points. For Chris, follow-up data points for LNF and LSF were both above all intervention data points, whereas data points for FSF showed a small drop in performance level. Sam's follow-up data showed similar levels of performance for LNF and LSF and an increase in FSF. Jill's follow-up data for LNF and LSF were quite inconsistent, making a comparison to end-of-intervention data points somewhat more difficult. For LNF, Jill had one data point at similar levels and one data point well below end-of-intervention data points. For LSF, Jill had one data point at similar levels and one data point well above end of intervention data points. Jill's FSF data were at similar levels of performance. For Robin, follow-up data points for LNF and FSF were both above all intervention data points, whereas data points for LSF were at similar levels. Tyler's follow-up data indicated similar levels of performance for both LNF and FSF and a substantial

drop in LSF as compared to end-of-intervention data points. In the next paragraph, data is presented in a consolidated format.

Follow-up data for LNF indicated that two students scored above end-of-intervention data points, three students scored at similar levels, and one student scored at mixed levels, with one data point at similar levels and one data point well below end-of-intervention data points. Thus, five of the six students scored at or above similar levels in follow-up as compared to end-of-intervention. Follow-up data for LSF indicated that one student scored above end-of-intervention data points, two students scored at similar levels, two students scored below end-of-intervention data points, and one student scored at mixed levels, with one data point at similar levels and one data point well above end-of-intervention data points. Three of six students scored at or above similar levels in follow-up as compared to end-of-intervention. Follow-up data for FSF indicated that two students scored above end-of-intervention data points, three students scored at similar levels, and one student scored below end-of-intervention data points. As with LNF data, five of the six students scored at or above similar levels in follow-up as compared to end-of-intervention.

PALS Data

Throughout the intervention on a schedule of approximately every twelfth session, I administered the PALS letter name and PALS letter sound assessments for each participant. The fourth progress monitoring session for Group B was skipped due to scheduling difficulties. Results were used as a formative assessment measure. Previously taught letter names and sounds that students did not know or had difficulty remembering were integrated into the alphabet review portion of the lesson. Tables 14 and 15 display results of the PALS assessments. On the letter name assessment, students identified a mean of 6.1 lowercase letter names. On

assessment sessions one through five during the intervention phase, students identified a mean of 16.7, 18.4, 20.3, 22.2, and 21.9 letter names respectively. A progressive increase is seen from pretest scores on monitoring sessions one through four.

On the letter sound assessment, students identified a mean of 5.2 letter sounds when presented with uppercase letters. On assessment sessions one through five during the intervention phase, students identified a mean of 12.4, 16.1, 18.0, 20.0, and 20.6 letter sounds respectively. A progressive increase is seen from pretest scores on all monitoring sessions.

Table 14. PALS Letter Name Data

Participant	LNPre	LN1	LN2	LN3	LN4	LN5
Ben	11	10	14	17	18	20
Chris	5	18	19	19	20	21
Sam	11	24	22	22	25	26
Jill	2	8	11	18		18
Robin	2	15	14	18		20
Tyler	4	11	16	19		22
Kevin	5	26	26	25	25	25
Tony	4	13	19	20	21	22
Frank	11	25	25	25	24	23
Mean	6.1	16.7	18.4	20.3	22.2	21.9

Table 15. PALS Letter Sound Data

Participant	LSPre	LS1	LS2	LS3	LS4	LS5
Ben	4	9	14	13	18	19
Chris	2	6	12	19	20	22
Sam	8	21	20	21	23	22
Jill	3	8	11	13		18
Robin	2	14	14	19		21
Tyler	4	7	14	18		18
Kevin	11	21	22	20	22	24
Tony	5	12	20	20	20	19
Frank	8	14	18	19	17	22
Mean	5.2	12.4	16.1	18	20	20.6

Inter-Assessor Agreement

As described in Chapter Three, inter-assessor agreement data was collected during the two-stage screening process, baseline, and intervention phases. Inter-assessor agreement was calculated using the formula Agreements / (Agreements + Disagreements) X 100%. According to Kennedy (2005), at least 80% agreement is a typical standard in applied research. Table 16 displays overall agreement for each measure administered during the two-stage screening process. Overall agreements ranged from 98.1% to 100%, indicating a high percentage of agreement for all measures.

Table 16. Prestest Inter-Assessor Agreement

Assessment	LNF	PALSLN	PALSLS	EL	BW	SM	RON
Total Agreement	99.1%	99.4%	98.1%	100%	100%	100%	99.5%

Note. LNF = Letter Naming Fluency; PALSLN = PALS-K Letter Naming; PALSLS = PALS-K Letter Sounds; EL = Elision; BW = Blending Words; SM = Sound Matching; RON = Rapid Object Naming.

Table 17 displays overall agreement for each of the three dependent measures administered during baseline and intervention phases. Overall agreements ranged from 95.1% to 98.8%, indicating a high percentage of agreement for all measures.

Table 17. Dependent Measures Inter-Assessor Agreement

Assessment	LNF	LSF	FSF
Total Agreement	98.8%	96.2%	95.1%

Note. LNF = Letter Naming Fluency; LSF = Letter Sound Fluency; FSF = First Sound Fluency

Fidelity of Implementation

As stated previously, fidelity data was collected for twenty-five percent of instructional sessions for each group. Treatment fidelity was calculated as the percent of components present out of the total number of components applicable for each lesson. Table 18 lists the percentage

of components present for each group for all observed sessions, as well as the combined treatment fidelity for all three intervention groups. Treatment fidelity for Groups A, B, and C were 96.9%, 98.5%, and 99% respectively. Fidelity percentages of individual sessions ranged from 84.6% to 100%. A fidelity percentage of 84.6% corresponded to two components not present of the 13 applicable components. Because of a transition from recess, some instructional time was lost from the 25-minute sessions for Group A. Based on the fidelity checklist, this lost instructional time is the primary reason for a lower fidelity percentage relative to groups B and C. Overall, treatment fidelity was 98.1% for all three groups combined.

Table 18. Treatment Fidelity Percentages

Group	Percentages
A	96.9%
B	98.5%
C	99.0%
Combined	98.1%

Summary

In summary, results indicated an effect of the independent variable on dependent variables at two of the three manipulations of the independent variable, diminishing evidence of a causal relationship. Differences in performance level and trend in the data were evident for five students on letter name knowledge, six students on letter sound knowledge, and six students on first sound identification skills. Inter-Assessor Agreement collected during the two-stage screening process and on dependent measures indicated a high percentage of agreement for all measures administered. Treatment fidelity percentages were high for all three groups, with an average of 98.1% of intervention components present during instruction.

CHAPTER FIVE: DISCUSSION

In this chapter, I first summarize and discuss the results of the early literacy intervention on the alphabet knowledge and phonemic awareness skills of study participants. Next, limitations of the study are discussed. Implications for research follow. Finally, implications for practice are discussed.

Summary and Discussion of Results

This study utilized a two-stage screening process to rank-order kindergarten students originally designated as at-risk according to DIBELS Next screening data administered in the beginning of the school year. The additional screening process began about two months into the school year and attempted to differentiate students with the most intensive literacy needs. Nine students who were identified as most at-risk were then selected to receive an intensive early literacy intervention. Students received instruction in three groups of three students each with the independent variable introduced in staggered fashion at three different points in time. At least three introductions of the independent variable are necessary to establish a cause and effect relationship between the independent and dependent variable (Barlow et al., 2009; Horner et al. (2005); Kratochwill et al. (2010).

According to the criteria identified by Kratochwill et al. (2010), this multiple baseline study meets evidence standards for single-case research. Specifically, the independent variable was systematically manipulated by the researcher at three different points in time, each baseline and intervention phase had a least five data points, outcome variables were measured systematically over time with inter-assessor agreement collected during at least 20% of observational sessions, and inter-assessor agreement met the minimum percentage of overall agreement of 80%. Additionally, fidelity of implementation was observed during 25% of

instructional sessions for each of the three groups, and overall, 98.1% of instructional components were present.

Kratochwill et al. (2010) also provided rules for demonstrating evidence of the cause and effect relationship between the independent variable and outcome variables. According to their criterion, at least three demonstrations of intervention effect are needed with no demonstrations of non-effects (although Kratochwill et al. stated that there is no formal basis for this recommendation). In the next two sections, this criterion is applied to the three outcome measures for this study. The terms time one, time two, and time three noted below correspond to the introduction of the independent variable for Group A, Group B, and Group C respectively.

Alphabet Knowledge

Visual analysis including CDC and NAP calculations revealed effects at both time one and time two for letter name knowledge as measured by LNF. However, no effect was discerned at time three, with one student showing a significant change in performance level in letter name knowledge just prior to introduction of the independent variable, and the other two students showing little or no growth from baseline to intervention phase.

For letter sound knowledge as measured by LSF, a similar pattern emerged. Visual analysis again revealed effects at both time one and time two, with no effect apparent at time three. In contrast to letter name knowledge, Group C results were somewhat mixed. Visual analysis revealed a clear change in performance level and trend for Kevin. The other two students in Group C, Tony and Frank, again exhibited little or no growth from baseline phase to intervention phase.

Throughout the intervention phase, additional progress monitoring data was collected using PALS letter name and letter sound progress monitoring tools. As described in Chapter Three, both measures are untimed and consist of exactly 26 items. Scores range from 0 to 26. As LNF and LSF do not indicate specifically what letter names and sounds students know and still need to work on, the PALS measures were used to further guide what letters and sounds to integrate into the review portion of the lessons. On the PALS letter name measure given at pretest, students averaged 6.1 letter names correct and scores ranged from 2 to 11 letter names. At the conclusion of the intervention phase, students averaged 21.9 letter names and scores ranged from 18 to 26 letter names (note: the PALS letter name measure consists of lowercase letters only, whereas LNF probes consist of uppercase and lowercase). On the PALS letter sound measures, students averaged 5.2 letter sounds correct and scores ranged from 2 to 11 letter sounds. At the conclusion of the intervention phase, students averaged 20.6 letter sounds and scores ranged from 18 to 24 letter names (note: the PALS letter sound measure consists of uppercase letters only, whereas LSF probes consist of lowercase only).

Keeping in mind differences in the use of uppercase and lowercase letters in the AIMSweb and PALS probes, the PALS measures indicated that, on average, students knew a vast majority of lowercase letter names and could produce the sound of a vast majority of uppercase letters. This is not to suggest that differences in skill level as measured by PALS alphabet probes can be attributed to the intervention, as these measures were only administered repeatedly during the intervention phase. The point being that despite this knowledge only some students, based on LNF and LSF data, appeared to transfer this knowledge to the automatic production of letter names and sounds. Recall that in their meta-analysis of alphabet learning, Piasta and Wagner (2010) found moderate effects on alphabet knowledge on untimed and timed

letter name and sound assessments collectively considered. When letter naming fluency assessments were parsed out, no effect was found for this particular measure. Perhaps this suggests an instructional focus at the level of automaticity in addition to frequent practice and review. Although frequent practice and review was a component of the intervention in this study, there was not a focus on automaticity.

Recall also that Simmons et al. (2007) found that 30 minutes of highly specified code-focused instruction produced significantly higher outcomes on decoding and word attack skills of kindergarteners than 15 minutes of highly specified instruction, particularly for those students with the lowest letter naming fluency scores. The approximate 16 minutes of code-focused instruction in this study was clearly not sufficient for some participants in this study. Specifically, Ben, Tony, and Frank made minimal or no gains on LNF, and Robin, Tony, and Frank made minimal or no gains on LSF data.

Another point of consideration is how the growth of participants in this study compared to grade-level norms. Even if data analysis of performance level and trend unequivocally indicated a cause and effect relationship between the independent variable and LNF and LSF data, it would be meaningful to know if student growth was sufficient to elevate students out of the at-risk category. Using the 30th percentile as the cut-off point for risk status as was used in the two-stage screening process, students would have needed winter and spring scores on LNF of at least 35 and 44 respectively. AIMSweb LSF winter and spring scores at the 30th percentile are 18 and 31 respectively. As student scores were collected at different points in the school year and at times between winter and spring benchmark periods, I can only estimate if students reached these levels. Looking at median scores of students' last three data points in the intervention phase, only one student was near an estimated 30th percentile for LNF at the time of

testing. Follow-up data indicated that one additional student exceeded the spring 30th percentile score. Looking at LSF data, three of the nine students were near or above the estimated 30th percentile score at the time of testing. This comparison suggests that rates of improvement of alphabet knowledge for the majority of participants in this study were insufficient to elevate them out of the risk category.

Phonemic Awareness

Visual analysis of phonemic awareness outcomes as measured by FSF revealed a rather complex pattern of results. Clear effects were seen for all three students in Group A, or at time one. A degree of variability existed in the data for all three students in Group B in the baseline phase, intervention phase, or both. Despite the inconsistency of the data, Jill and Tyler both exhibited an increase in performance level and trend from baseline to intervention phase, as well as meet the criteria for the CDC approach. Because of the variability and delayed effect in Tyler's data, it is difficult to attribute increases in level and trend of FSF to the independent variable. Robin's FSF data was perhaps the most contradictory and therefore integrating dimensions of within and between phase differences did not produce an unequivocal determination of effect. However, since a change in level and rate of improvement was evident after three weeks into the intervention phase and the CDC method showed all intervention data points above the mean and trend lines from baseline, I determined the intervention effective. Considered collectively, results for Group B support evidence of a cause and effect relationship at time two. At time three, results were again mixed, with only one student showing a clear increase in skill level from baseline to intervention phases.

In this paragraph, I compare the growth of participants' phonemic awareness skills to grade level norms. DIBELS Next, unlike AIMSweb, does not provide percentile rankings for

student scores for beginning, middle, and end of year testing. Additionally, FSF is not a standard assessment for end of year testing. Consequently, I can only make a comparison to the FSF mid-year cut point for risk, which is 42 correctly identified first sounds. At the time of final testing, four of the nine students were at or above the cut score using the median score of the last three data points of the intervention phase. As with alphabet knowledge, this comparison suggests that rates of improvement of phonemic awareness for the majority of participants in this study were insufficient to elevate them out of the risk category.

In summary, the integration of within and between phase data in this study revealed an effect of the independent variable on outcome measures for two points in time and no effect for one point in time. Consequently, the study does not meet the minimum standard of evidence as defined by Kratochwill et al. (2010) of three demonstrations of an effect, meaning that observed differences in outcomes for some students cannot be attributed to the independent variable. Individual analyses revealed that five of the nine participants in this study exhibited differences in level of performance and rate of improvement in LNF, six of nine in LSF, and six of nine in FSF.

Additionally, the intervention was not effective in elevating the majority of the most at-risk kindergarten students who participated in this study to levels above the 30th percentile on AIMSweb alphabet knowledge measures and above the DIBELS Next cut point for FSF. As reasons for less than desirable outcomes cannot be teased out, I can only speculate. The early literacy intervention in this study was intensified in the following respects: group sizes of three students, daily intervention sessions of 25 minutes each, and explicit or highly-specified instruction. It may be that more time per session was needed for code-focused components for participants in this study who had the lowest alphabet knowledge and phonemic awareness skill

level, a conclusion consistent with findings from Simmons et al. (2007). Extending the number of sessions from 60 to 100 or more may have also elevated students to higher levels of performance following intervention (see Simmons et al., 2011; and Lennon & Slesinski, 1999).

Limitations

Several limitations of this study should be noted. First, with a small sample size (n = 9), results of the study are not generalizable to other populations of at-risk kindergarten students. To be generalizable, this study needs to be replicated with other populations of students and implemented in different settings by multiple researchers and interventionists (Barlow et al., 2009; Birnbrauer, 1981; Kratochwill et al., 2010). Results can then be synthesized together, as well as synthesized with results of other similar studies. Before combining studies, Kratochwill et al. (2010) recommended a minimum of five high-quality studies conducted by at least three different research teams at three different settings. Second, the transportability, or extent to which the intervention could be implemented as designed by school practitioners, has not been explored.

A third limitation was the duration of the study. Because of the two-stage screening process and staggered starting and ending times, the intervention was limited to 60 instructional sessions. The intervention could be further intensified with a longer duration, perhaps at least 100 sessions, which corresponds to the number of sessions suggested by Simmons et al. (2011) and used to characterize early reading interventions as intensive by Wanzek & Vaughn (2007) in their synthesis. Consider also that Lennon and Slesinski (1999) found that 20 weeks of instruction was necessary for students with the lowest skill levels to improve their relative standing on outcome measures as compared to grade level peers.

A fourth limitation to be noted is the use of FSF as a measure of phonemic awareness. FSF was chosen instead of a phonemic segmentation fluency (PSF) task because first sound identification is an easier task in the developmental hierarchy of phonemic awareness skills than phonemic segmentation (see Paulson, 2004 and Pufpaff, 2009). Since the study involved the most at-risk students with the intention of intervening early in the school year, I believed FSF would be the more sensitive and appropriate measure. As the study progressed, this measure became somewhat problematic for two reasons. First, some students did not begin receiving intervention services until the mid-point of the school year. At this point in the school year, PSF may have been the more appropriate measure. Second, the DIBELS Next FSF probes consist of 30 items. As students were responding more quickly to task items over time, some students were finishing all 30 items in less than one minute, thus creating a ceiling effect.

Last, the meaning-focused components of the intervention were not measured in baseline and intervention phases. At present, reliable and valid measures of alphabet knowledge and phonemic awareness exist. Reliable, efficient, and repeated methods for measuring student growth in areas such as vocabulary and oral language skills are needed in order to include these constructs as outcome measures in single-case research.

Implications for Research

In this study, a two-stage screening process was designed and implemented with a two-pronged purpose: earlier identification of kindergarten students most at-risk for future reading failure and earlier implementation of intensive supports. As has been suggested with older students by Vaughn et al. (2010), the intention here was to bypass less intensive supports and move directly to more intensive supports for students with the most intensive needs. Measures selected for the two-stage screening were based on prior research investigating the best early

predictors of future reading success for kindergarten students (Catts et al., 2001; Catts et al., 2013; Hagan-Burke et al., 2011; Ortiz et al., 2012; Parrila et al., 2009; Schatschneider, et al., 2004; Simmons et al., 2013; Torgesen et al., 1999; Vellutino et al., 2008). Measures were selected from the following three categories: alphabet knowledge, phonological awareness, and rapid automatic naming. These measures most closely resemble the measures found in Catts et al. (2013) to be the best predictors of future reading outcomes. Results were then combined together by creating *z* scores from student raw scores. Both untimed and timed measures of alphabet knowledge were used in the screening process. Consequently, alphabet knowledge measures were weighted more heavily in the *z* score equation (i.e. 50% of the total summed *z* score). Future research is needed to validate this process and substantiate this method for combining and weighting results from a variety of early literacy measures in order to most accurately identify those most at-risk.

As noted previously, the independent variable was researcher designed and implemented for the first time in this study. To further explore the effects of this intervention on literacy outcomes of students with the lowest skill levels at kindergarten entry, future research involving direct replication is needed. With only one study consisting of nine participants, results are not generalizable to other populations of at-risk kindergarten students. Additionally, systematic replication could be used to extend initial research findings. According to Kennedy (2005), systematic replication involves changing a specific aspect of the research design to analyze the effect of the independent variable on the dependent variable. Based on the results of the alphabet knowledge measures, one such replication could involve extending session length for code-focused components to about 25 minutes per session. This timeframe approximates suggestions and findings from Simmons et al. (2007) and Torgesen (1999).

A review of the literature indicated that additional studies are needed to explore the effects of alphabet knowledge instruction for kindergarten students designated as at risk. The results on alphabet knowledge outcomes for participants in this study support this assertion. Future studies are needed to more systematically explore different approaches and emphases of alphabet instruction for students at risk, such as including or not including uppercase letters, teaching uppercase and lowercase letters together or apart, investigating the order of alphabet sequence, and investigating relative time spent on teaching letter names, sounds, and formation.

Future studies could also investigate the use of different alphabet knowledge dependent measures. Both the LNF and LSF probes consist of ten rows of ten letters in each row. For students who are successfully producing only a few correct letter names and sounds per minute, probes with larger font and fewer letters per page may be more reliable and sensitive measures of student knowledge and growth. From my observations, some of the participants in this study had difficulty engaging with these probes. Another potential issue with the LNF subtest is the inclusion of all 26 uppercase and lowercase letters. Consider that in this intervention, two uppercase and two lowercase letters were introduced in three-day periods. With this schedule, it takes about eight weeks to initially present all 52 letters. Since all 52 letters are presented in scrambled order across LNF probes, the true extent of student learning may be obscured. This may in part explain the relatively gradual rates of improvement seen in this study and the "no effect" calculated for measures of LNF in the meta-analysis conducted by Piasta and Wagner (2010). More sensitive measures to determine student response to intervention would enable practitioners to make more responsive instructional decisions.

Two interesting design elements of the intervention in this study were the use of word families and word chains during initial decoding instruction to provide additional elements of

support. For example, to facilitate the integration of alphabet knowledge and phonemic awareness, words chosen for a single session all came from the same common word family, such as "–at." Consequently, after students decoded the first word of the session (e.g. "sat"), only the first letter of the word was changed to create the next word. By repeating this process, the second and third letters remained constant. This created a great deal of repetition on frequent letter sounds and word patterns. To increase the challenge and allow for additional word patterns beyond the most common word families, word chains eventually replaced the word families. With a word chain, one letter is again replaced from one word to the next, but the change now involves the first, middle, or last letter of the word, thereby reducing the predictability of word families but still retaining an element of repetition. Systematic replication could further explore the effect of these elements by including additional dependent measures such as a test of nonsense word fluency or a word fluency test consisting of words with common word families.

Implications for Practice

A number of implications for practice can be drawn from this study. One important consideration is the practical utility and feasibility of a two-stage screening process designed to differentiate more-difficult-to-remediate students from less-difficult-to-remediate students. In the school in which this study took place, 38 of 80 students, or 48%, were designated as below benchmark or well below benchmark using fall benchmark scores from DIBELS Next. About eight weeks following this initial screening, a second screening process consisting of two stages was implemented. The purpose of the first stage was to reassess those students originally designated as at-risk by DIBELS Next. The first stage was intended to be relatively efficient and consisted of one alphabet knowledge measure and one phonemic awareness measure. Students who scored below the 30th percentile on either measure were administered a larger screening

battery as part of stage two. Of the 25 students who participated in stage one, only four scored above the 30th percentile on both measures. Consequently, 84% of the students who went through the first stage in the screening process participated in the stage two screening, which lasted about 25-30 minutes per child. For more feasible implementation, a lower percentile criterion may be necessary to reduce the percentage of students who take the larger assessment battery while still accurately identifying those students most at risk. For example, Simmons et al. (2013) reported that students who scored below the 16th percentile on either the Rapid Object Naming subtest of the CTOPP or the letter identification subtest from the Woodcock Reading Mastery Test-Revised/Normative Update were more likely to need intervention services than students scoring above the 16th percentile.

A second implication pertains to the timeframe of implementing intensive intervention services. In the context of this study, intervention services began for Group A in the thirteenth week of the school year. Theoretically, if the first screening was conducted during the second week of school, the two-stage screening process could take place during the eighth and ninth weeks. Intensive and less intensive intervention services could begin approximately during the tenth week of school for those students identified as at risk. This schedule would allow for about twenty-six weeks of intervention services. For the students in this study, additional time for services appeared to be needed. Results from prior studies suggest the same for some students (Al Otaiba & Fuchs, 2006; Lennon and Slesinski, 1999; Musti-Rao & Cartledge, 2007; O'Connor, 2000; Scanlon et al., 2005). In the context of a Response to Intervention framework, those students receiving more and less intensive intervention services would be continuously monitored for progress. Based on progress monitoring data, students would be moved into

different groups based on individual response (see Lennon & Slesinski, 1999 and Simmons et al., 2011 for similar suggestions regarding regrouping based on student response).

A third implication for practice pertains to continued monitoring of student progress. In this study, follow-up data was collected within a limited timeframe following intervention and for Group A and B participants only because of time limitations. With this in mind, phonemic awareness follow-up data exhibited that five of the six students generally maintained their end-of-intervention performance level, with only one student showing a dip in performance level. Alphabet knowledge follow-up data, particularly letter sound fluency, revealed a more complex picture, with some students appearing to make continued growth, others maintaining similar performance levels, and others showing skill regression. At least for the students in this study, continued monitoring of progress, especially of alphabet knowledge, appears necessary to determine the continued need for intervention services. The need for continued monitoring is supported from findings of the two-year study conducted by Vellutino et al. (2006) that spanned kindergarten and first grade. In their study, follow-up testing in third grade indicated that some students who were no longer at risk or less difficult to remediate at the end of kindergarten and first grade were again at risk.

A fourth implication for practice pertains to the coordination of instructional services. In this study, all participants received some whole group literacy instruction from Reading Street and Zoo Phonics (Safari Learning, 2007) with their classroom teacher, small group literacy instruction from Reading Mastery, and the small group instruction I delivered. Based on a pre-established scope and sequence, the code- and meaning-focused instructional components of this study did not necessarily align with other whole group and small group instruction students were receiving, and because of staggered start times, aligned differently for each group. The sight

words selected for instruction in this study did correspond generally to words students were learning in their core instruction, but were not necessarily in temporal alignment. In the context of multi-tiered systems of support, further coordination of services than occurred in this study may be needed.

One final implication for practice is the issue of student attendance. As participants received instruction in small groups as opposed to one on one, sessions were still held despite student absences, which is of course consistent within typical school practice. Consequently, not all students attended every session. Specifically, attendance rates ranged from 77% to 100%, with only one student attending every session. However, six of the participants attended 97% to 100% of the sessions, or between 58 to 60 sessions. Ben had the lowest rate of attendance at 77% (i.e. 46 sessions), due to frequent absences from school. Because of a scheduling conflict, Ben also left about half way through eight, or 13%, of the instructional sessions. Attendance rates for Tyler and Tony were both 83% (i.e. 50 sessions). Tyler missed sessions both because of absences and behavioral reasons. Tony's low rate of attendance was due to frequent absences.

It is obviously difficult to know the impact on outcome measures of missed sessions. The instructional sessions in this study built upon and were connected to prior sessions. Additionally, participants who had frequently missed sessions typically missed whole-group and targeted small-group literacy instruction as well, which could likely have an additional impact on outcome measures. Perhaps additional family supports were needed to increase the attendance rate for these three students.

Final Thoughts

In this study, intervention supports were intensified for kindergarten students identified as most at risk for future reading failure. Existing evidence indicates that early literacy intervention is effective for most students, but that some students do not respond adequately to intervention supports. In a multi-tiered instructional framework, instruction is delivered with differing levels of intensity. For kindergarten students, typical models are structured so that all students identified as at risk receive intervention supports at Tier Two, or supplemental levels of support. If students do not respond to instruction, intervention supports are intensified, but perhaps not until first grade. Why wait? To further reduce the number of students who do make adequate progress, differentiated levels of support can be delivered to at-risk students in kindergarten based on need as indicated by assessment data.

Any screening battery will likely identify students who appear to need supports that in actuality do not, and on the flip side, fail to identify some students who do need intervention supports, or who need supports more intensively than indicated. In multi-tiered frameworks, all students should be regularly screened and student progress should be monitored regularly for all students at risk, regardless of their degree of risk. As school-wide assessment data is being collected and student response to intervention is being measured, school personnel must also be responsive, continuously making decisions as to what levels of support to deliver to which students and for how long. To elevate the greatest number of at-risk students to levels of proficiency, the highest possible quality of instruction is needed. More research is needed to identify the type and intensity level of early literacy instruction necessary to successfully elevate students most at risk to a level of proficient reading.

References

AIMSweb (2012). *Aimsweb technical manual*. Bloomington, MN: Pearson, Inc. Retrieved from: http://www.aimsweb.com/products/features/assessments/reading-cbm

Al Otaiba, S., Connor, C. M., Folsom, J. S., Greulich, L., Meadows, J., & Li, Z. (2011). Assessment data-informed guidance to individualize kindergarten reading instruction: Findings from a cluster-randomized control field trial. *The Elementary School Journal, 111*(4), 535-560.

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*(5), 300-316.

Al Otaiba, S., & Fuchs, D. (2006). Who are the young children for whom best practices in reading are ineffective? *Journal of Learning Disabilities, 39*(5), 414-431.

Al Otaiba, S., Schatschneider, C., & Silverman, E. (2005). Tutor-assisted intensive learning strategies in kindergarten: How much is enough? *Exceptionality, A Special Education Journal, 13*(4), 195-208.

Barlow, D. H., Nock, M. K., Hersen, M. (2009). *Single Case Experimental Design: Strategies for Studying Behavior Change*. Boston, MA: Pearson Education, Inc.

Barnes, A. C., & Harlacher, J. E. (2008). Clearing the confusion: Response-to-Intervention as a Set of Principles. *Education and Treatment of Children, 31*(3), 2008.

Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal, 107*(3), 251-271.

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology, 98*(1), 44-62.

Birnbrauer, J. S. (1981). External validity and experimental investigation of individual behaviour. *Analysis and Intervention in Developmental Disabilities, 1*, 117-132.

Bradley, M. C., Daley, T., Levin, M., O'Reilly, F., Parsad, A., Robertson, A., & Werner, A. (2011). *IDEA national assessment implementation study: A final report* (NCEE 2011-4027). Washington, DC: Institute of Educational Sciences. Retrieved from http://ies.ed.gov/ncee/pubs/20114026/pdf/20114027.pdf

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: a research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools, 32*, 38-50.

Cavanaugh, C. L., Kim, A., Wanzek, J., & Vaughn, S. (2004). Kindergarten reading interventions for at-risk students: Twenty Years of Research. *Learning Disabilities: A Contemporary Journal, 2*(1), 9-21.

Chard, D. J. (2012). A glass half full: A commentary on the special issue. *Journal of Learning Disabilities, 45*(3), 270-273.

Compton, D. L., Gilbert, J. K., Jenkins, J. R., Fuchs, D., Fuchs, L. S., Cho, E., … & Bouton, B. (2012). Accelerating chronically unresponsive children to Tier 3 instruction: What level of

data is necessary to ensure selection accuracy? *Journal of Learning Disabilities, 45*(3), 204-216.

Coyne, M. D., Little, M., Rawlinson, D., Simmons, D., Kkwok, O., Kim, M., Simmons, L., et al. (2013). Replicating the impact of a supplemental beginning reading intervention: The role of instructional context. *Journal of Research on Educational Effectiveness, 6*(1), 1-23.

Coyne, M. D., McCoach, D. B., & Kapp S. (2007). Vocabulary intervention for kindergarten students: Comparing extended instruction to embedded instruction and incidental exposure. *Learning Disabilities Quarterly, 30*, 74-88.

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli, R. Jr., & Kapp, S. (2009). Direct vocabulary instruction in kindergarten: Teaching for breadth versus depth. *The Elementary School Journal, 110*(1), 1-18.

Coyne, M. D., Simmons, D. C., Kame'enui, E. J., & Stoolmiller, M. (2004). Teaching vocabulary during shared storybook readings: An examination of differential effects. *Exceptionality, 12*(3), 145-162.

Evans, M. A., Bell, M., Shaw, D., Moretti, S., & Page, J. (2006). Letter names, letter sounds and phonological awareness: An examination of kindergarten children across letters and of letters across children. *Reading and Writing, 19*(9), 959-989.

Fisher, W. W., Kelley, M. E., Lomas, J. E. (2003). Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. *Journal of Applied Behavior Analysis, 36*(3), 387-406.

Flynn, K. S. (2011). Developing children's oral language skills through dialogic reading: Guidelines for implementation. *Teaching Exception Children, 44*(2), 8-16.

Fuchs, D., Compton, D. L., Fuchs, L. S., Bryant, V. J., Hamlett, C. L., & Lambert W. (2012). First-grade cognitive abilities as long-term predictors of reading comprehension and disability status. *Journal of Learning Disabilities, 45*(3), 217-231.

Fuchs, D., & Fuchs L. S. (2006). Introduction to Response to Intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*(1), 93-99.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2010). Rethinking Response to Intervention at middle and high school. *School Psychology Review, 39*(1), 22-28.

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-Intervention: A decade later. *Journal of Learning Disabilities, 45*(3), 195-203.

Gay L. R., Mills, G. E. & Airasian, P. (2012). *Educational research: Competencies for analysis and applications.* Boston, MA: Pearson Education, Inc.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2009). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades. A Practice Guide.* (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Services, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publication/practiceguides/

Good, R. H., & Kaminski, R. A. (2011a). *DIBELS Next assessment manual*. Eugene, OR: Dynamic Measurement Group. Retrieved from https://dibels.org/next/index.org/

Good, R. H., & Kaminski, R. A. (2011b). *DIBELS Next technical manual*. Eugene, OR: Dynamic Measurement Group. Retrieved from https://dibels.org/next/index.php/

Gormley, S., & Ruhl, K. L. (2005). Dialogic shared storybook reading: An instructional technique for use with young students in inclusive settings. *Reading & Writing Quarterly, 21*(3), 307-313.

Gyovai, L. K., Cartledge, G., Kourea, L., Yurick, A., & Gibson, L. (2009). Early reading intervention: Responding to the learning needs of young at-risk English language learner. *Learning Disability Quarterly, 32*(3), 143-162.

Hagan-Burke, S., Coyne, M., Kwok, O., Simmons, D. C., Kim, M., Simmons, L. E., … Ruby, M. M. (2011). The effects and interactions of student, teacher, and setting variables on reading outcomes for kindergartners receiving supplemental reading intervention. *Journal of Learning Disabilities, 46*(3), 260-277.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165-179.

Ikeda, M. J. (2012). Policy and practice consideration for Response to Intervention: Reflections and commentary. *Journal of learning Disabilities, 45*(3), 274-277.

Jones, C. D., Clark, S. K., & Reutzel, D. R. (2013). Enhancing alphabet knowledge instruction: Research implications and practical strategies for early childhood educators. *Early Childhood Education Journal, 41*(2), 81-89.

Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergarteners. *Language, Speech, and Hearing Services in Schools, 36*, 17-32.

Kennedy, C. H. (2005). *Single-case designs for educational research*. Boston, MA: Pearson Education, Inc.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc_scd.pdf

Invernizzi, M., Juel, C., Swank, L., & Meier, J. (2013). *Phonological awareness literacy screening K technical reference*. Charlottesville, VA: The Rector and the Board of Visitors.

Lennon, J. E., & Slesinski, C (1999). Early Intervention in Reading: Results of a screening and intervention program for kindergarten students. *School Psychology Review, 28*(3), 353-364.

Lyon, G. R., Shaywitz, S. E., Shaywitz, A., & Chhabra, V. (2005). Evidence-based reading policy in the United States: How scientific research informs instructional practices. *Brookings Papers on Education Policy,8*, 209-250.

Maggin, D. M., Swaminathan, H., Rogers, H. J., O'Keefe, B. V., Sugai, G., & Horner, R. H. (2011). A generalized least squares regression approach for computing effect sizes in single-case research: Application examples. *Journal of School Psychology, 49*, 301-321.

McBride-Chang, C. (1999). The ABCs of the ABCs: The development of letter-name and letter-sound knowledge. *Merrill-Palmer Quarterly, 45*(2), 285-308.

Moats, L. C. (2010). *Speech to print: Language essentials for teachers* (2nd ed.). Baltimore, MD: Paul H. Brookes Publishing Co., Inc.

Morgan, P. L. & Meier, C. R. (2008). Dialogic reading's potential to improve children's emergent literacy skills and behavior. *Preventing School Failure, 52*(4), 11-16.

Musti-Rao, S., & Cartledge, G. (2007). Effects of a supplemental early reading intervention with at-risk urban learners. *Topics in Early Childhood Special Education, 27*(2), 70-85.

National Early Literacy Panel. (2008). Developing early literacy: Report of the National Early Literacy Panel. Washington, DC: National Institute for Literacy. Available at http://www.nifl.gov/earlychildhood/NELP/ NELPreport.html National Reading Panel

Nelson, J. R., Benner, G. J., & Gonzalez, J. (2003). Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta-analytic review. *Learning Disabilities Research & Practice, 18*(4), 255-267.

O'Connor, R. (2000). Increasing the intensity of intervention in kindergarten and first grade. *Learning Disabilities Research and Practice,15*(1), 43-54.

Ortiz, M., Folsom, J. S., Al Otaiba, S., Greulich, L., Thomas-Tate, S., & Connor, C. M. (2012). The componential model of reading: Predicting first grade reading performance of culturally diverse students from ecological, psychological, and cognitive factors assessed at kindergarten entry. *Journal of Learning Disabilities, 45*(5), 406-417.

Oudeans, M. K. (2003). Integration of letter-sound correspondences and phonological awareness skills of blending and segmenting: A pilot study examining the effects of instructional

sequence on word reading for kindergarten children with low phonological awareness. *Learning Disability Quarterly, 26*, 258-280.

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy, 40*, 357-367.

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A Review of nine nonoverlap techniques. *Behavior Modification, 20*(10), 1-20.

Parrila, R., Kirby, J. R., & McQuarrie, L. (2009). Articulation rate, naming speed, verbal short-term memory, and phonological awareness: Longitudinal predictors of early reading development. *Scientific Studies of Reading, 8*(1), 3-26.

Paulson, L. H. (2004). *The development of phonological awareness skills in preschool children: From syllables to phonemes.* (Order No. 3166292, University of Montana). *ProQuest Dissertations and Theses,* 127 p. Retrieved from http://search.proquest.com/docview/305160145?accountid=14593. (305160145)

Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew Effect? *Journal of Educational Pyschology, 94*(1), 23-33.

Piasta, S. B., & Wagner, R. K. (2010). Developing early literacy skills: A meta-analysis of alphabet learning and instruction. *Reading Research Quarterly, 45*(1), 8-38.

Pufpaff, L. A. (2009). A developmental continuum of phonological sensitivity skills. *Psychology in the Schools, 46*(7), 679-691.

Salvia, J., Ysseldyke, J. E., & Bolt, S. (2007). *Assessment in special and inclusive education* (11th ed.). Belmont, CA: Wadsworth, Cengage Learning.

Samanich, T. T. (2004). *The effectiveness of the scott-foresman early reading intervention program on improvement of phonemic awareness and decoding skills for a sample of at-risk kindergarten students.* (Order No. AAI3124391, *Dissertation Abstracts International Section A: Humanities and Social Sciences,* , 831. Retrieved from http://search.proquest.com/docview/620639222?accountid=14593. (620639222; 2004-99017-007)

Scanlon, D. M., Vellutino, F. R., Small, S. G., Fanuele, D. P., & Sweeney, J. M. (2005). Severe Reading Difficulties – Can they be prevented? A comparison of prevention and intervention approaches. *Exceptionality, 13*(4), 209-227.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96*(2), 265-282.

Shinn, M. R., & Germann, G. (2006). *AIMSweb*. Eden Prairie, MN: Edformation.

Simmons, D. C., Coyne, M. D., Hagan-Burke, S., Kwok, O., Simmons, L., Johnson, C., … Crevecoeur, Y. C. (2011). Effects of supplemental reading interventions in authentic contexts: A comparison of kindergarteners' response. *Exceptional Children, 77*(2), 207-228.

Simmons, D. C., Kame'enui, E. J., Harn, B., Coyne, M. D., Stoolmiller, M., Santoro, L. E., … Kaufman, N. K. (2007). Attributes of effective and efficient kindergarten reading intervention: An examination of instructional time and design specificity. *Journal of Learning Disabilities, 40*(4), 331-347.

Simmons, D. C., Kim, M., Kwok, O., Coyne, M., Simmons, L., Oslund, E., … Rawlinson, D. (2013). Examining the effects of linking student performance and progression in a Tier 2 kindergarten reading intervention. *Journal of Learning Disabilities, XX*(X), 1-16.

Simmons, D. C., Taylor, A. B., Oslund, E. L., Simmons, L. E., Coyne, M., Little, M. E., … Kim, M. (2013). Predictors of at-risk kindergarteners' later reading difficulty: examining learner-by-intervention interactions. *Reading & Writing, 27*(3), 451-479.

Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review, 6*, 1-26.

Snow, C. E., Burns, S. M., & Griffin, P. (Eds.) (1998). *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.

Steinberg, W. J. (2011). *Statistics alive!* (2nd ed.). Los Angeles, CA: Sage Publishing, Inc.

Swoboda, C. M., Kratochwill, T. R., & Levin, J. R. (2010). *Conservative dual-criterion method for single-case research: A guide for visual analysis of AB, ABAB, and multiple-baseline designs* (WCER Working Paper No. 2010-13). Retrieved from University of Wisconsin–Madison, Wisconsin Center for Education Research website: http://www.wcer.wisc.edu/publications/workingPapers/papers.php

Torgesen, J. K. (2000). Individual differences in response to early intervention in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice, 15*(1), 55-64.

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. K. S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities:

Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities, 34*(1), 33-58.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., & Garvan, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91*(4), 579-593.

Trieman, R., Tincoff, R., Rodriguez, K., Mouzaki, A., & Francis, D. J. (1998). The foundations of literacy: Learning the sounds of letters. *Child Development, 69*(5), 1524-1540.

Trivette, C. M., Simkus, A., Dunst, C. J., & Hamby, D. W. (2012). Repeated book reading and preschoolers' early literacy development. *CELL Reviews, 5*(5), 1-13.

Turnbull, K. L. P., Bowles, R. P., Skibbe, L. E., Justice, L. M., & Wiggins, A. K. (2010). Theoretical explanations for preschoolers' lowercase alphabet knowledge. *Journal of Speech, Language, and Hearing Research, 53* 1757-1768.

Vaughn, S., Denton, C. A., & Fletcher, J. M. (2010). Why intensive interventions are necessary for students with severe reading difficulties. *Psychology in the Schools, 47*(5), 432-444.

Vaughn, S., & Fuchs, L. S. (2006). A response to "Competing views: A dialogue on Response to Intervention": Why Response to Intervention in necessary but not sufficient for identifying students with learning disabilities. *Assessment for Effective Intervention, 32*(1), 58-61.

Vaughn, S., Fletcher, J. M., Francis, D. J., Denton, C. A., Wanzek, J., Wexler, J., … Romain, M. A. (2008). Response to Intervention with older students with reading difficulties. *Learning and Individual Differences, 18*(3), 338-345.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Chen, R., Pratt, A., Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88*(4), 601-638.

Vellutino, F. R., Scanlon, D. M., Small, S., & Fanuele, D. P. (2006). Response to Intervention as a vehicle for distinguishing between children with and without reading disabilities: Evidence for the role of kindergarten and first-grade interventions. *Journal of Learning Disabilities, 39*(2), 157-169.

Vellutino, F. R., Scanlon, D. M., Zhang, H., & Schatschneider, C. (2008). Using response to kindergarten and first grade intervention to identify children at-risk for long-term reading difficulties. *Reading and Writing, 21*, 437-480.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive test of phonological processing* (2nd Ed.). Austin, TX: PRO-ED, Inc.

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review, 36*(4), 541-561.

Wanzek, J., Wexler, J., Vaughn, S., & Ciullo, S. (2010). Reading interventions for struggling readers in the upper elementary grades: A synthesis of 20 years of research. *Reading and Writing, 23*(8), 889-912. doi:http://dx.doi.org/10.1007/s11145-009-9179-5

Whitehurst, G. J. (1992). Dialogic Reading: An effective way to read to preschoolers. Retrieved from http://www.readingrockets.org/article/400/?theme=print

Wolery, M., & Ezell, H. K. (1993). Subject descriptions and single-subject research. *Journal of Learning Disabilities, 26*(10), 642-647.

Wolery, M., Busick, M., Reichow, B., & Barton, E. E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *The Journal of Special Education, 44*(1), 18-28.

Appendix A: Books and Vocabulary Words for Shared-Book Reading Activities

Books	Extended Vocabulary Words	Embedded Vocabulary Words
Imogene's Antlers	Advice, Consulted, Glared	Announced, Wandered, Rare
Otis	Bawl, Startle, Discover	Unwind, Explode, Tugged
Shy Charles	Embarrassed, Shy, Success	Trembled, Pretend, Nervous
Mrs. Potter's Pig	Speck, Clutching, Peeked	Complained, Grunted, Gobbled
A Visitor for Bear	Visitors, Bright-Eyed, Unbelievable	Wailed, Commanded, Impossible
Buster	Favorite, Ignore, Familiar	Terrified, Lonely, Route
Nothing like a Puffin	Creature, Beak, Hatch	Marvelous, Scales, Propeller
The Scarecrow's Hat	Swap, Wool, Delighted	Walking Stick, Swat, Sigh
Too Many Toys	Stashed, Squirm, Filthy	Miniature, Hazard, Plop
The Kissing Hand	Gently, Cozy, Teased	Strange, Palm, Scamper
Goodbye Hello	Neighborhood, Perfect, Explore	Doorman, Secret, Aquarium
The Pout-Pout Fish	Pout, Impolite, Brilliant	Frown, Mope, Unattractive
My Lucky Day	Preparing, Delicious, Exhausted	Hauled, Growled, Terrific
Muncha! Muncha! Muncha!	Gobble, Nibble, Furious	Hoe, Blossom, Enormous
The Wolf's Chicken Stew	Craving, Scrumptious, Joyfully	Terrible, Prey, Critter

Appendix B: Intervention fidelity checklist

Lesson Components	Component is present	Component is not present	Not Applicable for this lesson
Alphabet instruction begins with a review of previously taught letters			
Letter names are being taught/practiced and students are saying the names aloud			
Letter sounds are being taught/practiced and students are saying the sounds aloud			
The shape and/or formation of the letters are being taught/practiced (up to week 9)			
Phonemic awareness instruction is evident (first or last sound isolation or phonemic blending/segmenting of 2-4 sound words) or 2-3 sounds words are being decoded/encoded with letter tiles			
A picture book is read aloud or activities for words chosen for extended vocabulary			
Vocabulary words are being taught/practiced			
Students are asked literal questions, open-ended questions, or engage in a retell of the story			
Explicit modeling of skills is evident			
Corrective feedback is evident			
Alphabet instruction is approx. 9-10 minutes			
Phonemic Awareness instruction is approx. 4-5 minutes (or in week 8, phonemic awareness is blended into decoding/encoding instruction 14-15 minutes total)			
Shared-book reading is approx. 7-9 minutes			
Beginning in week 9, sight words are being taught/practiced/read			
Beginning in week 9, 2-3 sound decodable words are being spelled with letter tiles or by writing the words			