

Understanding Effective Instruction: The Role of Content in First Grade Reading Instruction

by

Stephanie J. Guthrie

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

(Education and Psychology)

2013

Doctoral Committee:

Professor Frederick J. Morrison, Chair

Professor Joanne L. Carlisle

Associate Professor Kai S. Cortina

Associate Professor Priti Shah

Acknowledgements

The research collected for this dissertation was funded by many entities at the University of Michigan: The Office of the Vice President for Research; the Psychology Department; the College of Literature, Science, & the Arts; and the Combined Program in Education and Psychology.

I would like to thank my committee chair, Fred Morrison, for his support, guidance, and chocolate malts. Thank you also to all of my committee members: Joanne Carlisle, Kai Cortina, and Priti Shah. They have provided me with valuable resources and advice on this project and beyond.

This research would not have been possible without my undergraduate research assistants: Ann Olson, who designed my teacher questionnaires and wrote her honors thesis using a portion of the data collected; Callan Banach; Sam Feldman; Hannah Jacobson; Nomi Kornfeld; Michelle Morgan; Bethany Pester; and Nikki Reiner. I would also like to thank Kim Palombo for her assistance with creating my worksheet coding system.

Finally, thank you to my family and friends for all of their love and support over the past five years, and especially to Ryan and Hinson for keeping me company during this last one.

Table of Contents

Acknowledgements.....	ii
List of Tables.....	iv
List of Figures.....	vi
Abstract.....	vii
Chapter 1. Overview and Rationale.....	1
Chapter 2. Review of the Literature.....	6
Chapter 3. Method.....	22
Chapter 4. Analyses and Results.....	36
Chapter 5. Discussion.....	48
References.....	82

List of Tables

Table 1. Average observed instruction and recommended instruction from the fall and spring timepoints in each of the four Pathways/ISI codes, separated by student reading ability.....	60
Table 2. Correlations between Pathways/ISI codes and students' achievement in decoding, comprehension, and vocabulary in the fall and spring.....	61
Table 3. Model results using distance from recommendation (DFR) scores in each of the Pathways/ISI codes to predict students' growth in decoding over the school year.....	62
Table 4. Model results using distance from recommendation (DFR) scores in each of the Pathways/ISI codes to predict students' growth in comprehension over the school year.....	63
Table 5. Model results using the amount of observed time spent in each of the Pathways/ISI codes to predict students' growth in decoding over the school year.....	64
Table 6. Model results using the amount of observed time spent in each of the Pathways/ISI codes to predict students' growth in comprehension over the school year.....	65
Table 7. Average CLASS domain scores.....	66
Table 8. Correlations amongst CLASS domain and dimension scores.....	67
Table 9. Correlations between CLASS scores and instructional variables in the fall and spring.....	68
Table 10. Model results using the amount of observed time spent in each of the Pathways/ISI codes to predict students' growth in decoding over the school year, with CLASS Domain scores predicting students' growth on decoding.....	69
Table 11. Model results using the amount of observed time spent in each of the Pathways/ISI codes to predict students' growth in comprehension over the school year, with CLASS Domain scores predicting students' growth on comprehension.....	70
Table 12. Correlations between student achievement and book readability discrepancy scores and students' growth over the school year in reading skills.....	71
Table 13. Correlations between text-related data (book difficulty, total number of books read, and time spent reading books) and all four types of literacy instruction in the spring and fall.....	72
Table 14. Model results using the amount of time students spent reading books during the observation, as well the amount of time they spent engaging with each individual book, to predict students' growth in decoding over the school year.....	73

Table 15. Model results using the amount of time students spent reading books during the observation, as well the amount of time they spent engaging with each individual book, to predict students' growth in comprehension over the school year.....	74
Table 16. Teacher ratings of students' literacy skills correlated with students' concurrent performance on reading assessments, as well as both CLASS and Pathways/ISI observation codes.....	75
Table 17. Model results using the amount of observed time spent in each of the Pathways/ISI codes to predict students' growth in decoding over the school year, with teacher's years of experience teaching first grade predicting the degree of growth in decoding.....	76
Table 18. Model results using the amount of observed time spent in each of the Pathways/ISI codes to predict students' growth in comprehension over the school year, with teacher's years of experience teaching first grade predicting the degree of growth in comprehension.....	77

List of Figures

Figure 1. Difference between students' scores on decoding in the fall and the average difficulty level of books interacted with in the fall.....	78
Figure 2. Difference between students' scores on comprehension in the fall and the average difficulty level of books interacted with in the fall.....	79
Figure 3. Difference between students' scores on decoding in the spring and the average difficulty level of books interacted with in the fall.....	80
Figure 4. Difference between students' scores on comprehension in the spring and the average difficulty level of books interacted with in the fall.....	81

Abstract

The present study focused on a currently central issue in educational science, namely how to conceptualize and measure effective classroom instruction. Although several observational methods, which capture aspects of the classroom related to students' cognitive development, are available for researchers and school personnel to implement in studies of instructional quality, this study explored relations between two established classroom observation measures (CLASS and Pathways/ISI), that have a strong evidence base on growth of reading skills in first-grade students. Further, the study expanded the focus to include an additional component (content difficulty) of instruction that has been separately linked to students' development.

The 233 students enrolled in this study came from 17 different first grade classrooms across six schools. Students were assessed on decoding and comprehension skills in the first and last marking period of the school year through one-on-one standardized assessments. Classrooms were observed using an adapted live-coding version of the Pathways/ISI coding scheme as well as the CLASS. Observations lasted during approximately 2 hours of literacy instruction, and were conducted in the second and third marking periods of the school year. To measure instructional content, the titles of books students and teachers used were recorded and coding for readability using the Flesch-Kincaid scale. Teachers and parents also responded to questionnaires at the end of the school year.

Results of models using the amount of time spent in each of two Pathways/ISI domains (teacher-managed code-focused instruction and child-managed meaning-focused instruction) showed that the more time children spent in each of those two domains was significantly related to decoding and comprehension skills. However, comparing the distance from recommendation of naturally occurring instruction in these same domains found no significant links. Comparing

across classroom observation measures, correlational evidence indicated that the Pathways/ISI observation codes and the CLASS scores for each classroom were unrelated, which provides support for the hypothesis that these two measures are independent. The addition of CLASS scores into these models showed that CLASS scores were significantly predictive of students' outcome scores in decoding and comprehension. In general, the difficulty or readability of texts students interacted with was not linked with their reading outcomes when added to a model of classroom instruction; however, interacting with books was a significant predictor of higher reading achievement. The alignment between book difficulty level and students' reading achievement showed a significant linear relation in the spring, but not in the fall. At both time points students were, on average, interacting with books within half a grade level of their current reading skills. A greater degree of challenge (e.g., students who read books leveled above their current reading ability) was significantly related to greater gains in student achievement over the school year.

In general, a number of limitations limited the power of the study to detect significant effects; however, this study again found that the two systems provide independent information about instruction. Thus, future studies using both CLASS and Pathways/ISI may be fruitful. Findings about book use and links between number of books read and students' skills at the end of the school year indicate that further work with text difficulty and students' interaction with books is needed.

Overall, this study's findings are an important first step in creating a more comprehensive dyadic measure of the aspects of classroom literacy instruction that are most effective at improving students' reading skills throughout the school year.

Chapter 1. Rationale and Overview

The current focus of numerous policy-makers, from government representatives to school officials, is improving the quality of education and specifically the effectiveness of schooling experiences for all students. Many policy and research-based initiatives have emphasized the importance of highly qualified and effective teachers to student achievement and success. These initiatives emerge as a result of evidence that schooling experiences differ greatly across and between students (e.g., Goodlad, 1984; Peterson & Fennema, 1985; Sanders, Wright, & Horn, 1997). Moreover, because students often transition to school with different cognitive, language, and social skills, these different experiences during school can exacerbate deficiencies in child knowledge and skill. The intention of these policies is to narrow the differences in student skills through high quality instruction.

Although many policy initiatives have targeted early learning experiences such as preschool and prekindergarten programs (e.g., More at Four, HeadStart), evidence from longitudinal investigations of student achievement has demonstrated the importance of each year of schooling to academic outcomes (Alexander, Entwisle and Olson, 2001; Guthrie, Connor, & Morrison, 2013). Indeed, a recent study of student achievement over multiple years of elementary school revealed differences between students who received three versus four years of highly effective instruction, with the effects of instruction cumulating over all four years (e.g., larger differences between students who received one versus four years of highly effective instruction) (Guthrie, Connor, & Morrison, 2013). Thus, an understanding and emphasis on high quality schooling experiences in every year of schooling is important to achieving the goal of improving student skills and knowledge.

The Importance of Schooling

In large-scale studies of state-wide testing programs, using multi-level analyses to tease apart sources of variance, classroom experiences routinely constitute a major if not the greatest unique contributor to academic achievement (Nye et al., 2004). Complementary findings from natural experiments reinforce the unique importance of schooling-related factors in shaping early literacy trajectories (Morrison et al., 2005). In a series of studies, Morrison and colleagues (Bisanz, Morrison et al., 1995; Christian, Morrison et al., 2000; Morrison, Smith, et al., 1995) compared students' growth on a variety of literacy, mathematics, and socioemotional skills between those students who just made versus missed the cutoff for school. Strong evidence for the unique impact of schooling related-processes on some but not all of the targeted skills was found repeatedly in this series of studies.

Specifically, Morrison and colleagues (1995) found that children who just made the cutoff for first grade showed stronger growth in phonemic segmentation skills but not syllabic segmentation compared to students who just missed the cutoff. This finding indicates that some aspects of children's language and literacy skills may be more sensitive to schooling, and provides evidence that schooling has a unique effect on various aspects of children's cognitive development. In a more recent study, Skibbe and colleagues (2008) conducted a cutoff study (comparing the skills of students who just made versus missed the cutoff for school entry) which utilized observations of classroom instructional activities to link students' ability with their exposure to schooling. This study found that differences in student ability between the two groups were predicted by the amount, type and content of instruction received by the first graders compared to their almost same-age kindergarten peers.

Finally, recent intervention work has demonstrated further potential for schooling to influence students' cognitive skills and academic achievement (Borman, Slavin et al., 2005; Connor et al., 2007; Sadoski & Willson, 2006). Connor et al. (2007; 2008) conducted a randomized control trial in elementary schools which demonstrated the positive effects of a first grade literacy intervention. This intervention used students' incoming literacy knowledge in order to design individualized student literacy instruction in public school classrooms using teacher-accessed computer software.

The pattern of findings across descriptive studies, natural experiments and interventions converges inexorably on the important role of classroom experiences as critical contributors to children's growing academic skills (Connor et al., 2007). The cumulative evidence suggests that instructional and related classroom practices (e.g., classroom management, teacher warmth/responsivity) shape students' academic achievement growth. Moreover, by accounting for the unique aspects of individual students' schooling experiences, it is becoming possible to characterize aspects of the classroom setting that account for variability in children's academic outcomes. The present study sought to measure and assess the impact of a subset of classroom experiences on first-grade children's reading growth

The Importance of Literacy Achievement

Reading skills are major focus during the early years of schooling, and are critical to educational success with links being made between reading ability and poverty and unemployment (NCES, 2009). However, a large number of American primary and secondary school students are unable to read at grade-level, and are often promoted to the next grade with little to no remediation (Bachman, Connor, & Morrison, 2005). Early literacy skills are influenced by a number of factors, including aspects of the home and community environment as

well as parental characteristics such as education (NICHD-ECCRN, 2004). One major contributor to growth in early literacy skills are early schooling experiences (NICHD-ECCRN, 2002b; Pianta, La Paro, Payne, Cox, & Bradley, 2002; Entwisle, Alexander, & Olson, 2006), which explains current research and policy emphasis on introducing critical reading skills at early ages in an effort to reduce gaps in literacy knowledge as a result of early environmental differences. Indeed, many have noted the shift in emphasis in preschool and kindergarten classrooms from social- and school-related skills to emergent and early literacy and mathematics skills as a result of this political focus on early education (Foster & Miller, 2007).

Given the importance of early experiences and skills to later literacy development, and the clear ramifications of having effective instruction for every student, it is still unclear what processes underlie effective or ineffective instruction. To this end, researchers have conducted multiple studies of classroom instruction as it predicts student outcomes posing the question: What is effective literacy instruction? As a result of these investigations, researchers are learning more about the types of instructional activities that benefit students of all achievement levels, specifically the importance of phonics- and meaning-related instruction and the role of the teacher's presence and assistance in facilitating learning.

Identifying Effective Instruction

Over the past decade, researchers have developed a number of methods for observing and classifying classroom instructional practices and activities. These observational methods can be identified broadly as either content specific or content general; that is, either targeting practices related to one content area (e.g., Protocol for Language Arts Teaching Observation, Mathematics Quality of Instruction, Quality of Science Teaching), or targeting global classroom practices and climate (e.g., teacher-student relationships, classroom management) (e.g.,

Classroom Assessment Scoring System). Use of these observational methods in studies of classroom quality and student achievement has linked specific processes and characteristics to improvement in student knowledge and skills. Recent investigations by the Gates Foundation as part of the MET project attempted to assess the impact of several classroom observational systems on student achievement. They found that direct measures of teaching effectiveness (value-added) were consistently the best predictor of teachers' ability to improve student achievement scores (Kane & Staiger, 2012). In addition, the researchers noted that each classroom observation was related to student performance, but that in order to achieve a reliable measure of teacher performance multiple classroom observation tools must be used and combined for each teacher across a range of instructional activities. A similar study of earlier grades is decidedly lacking, especially given the importance of early instruction and cognitive skill development. The present study compared the predictive power of two evidence-based observational systems on reading development in a sample of first-grade students.

Specifically, two measures of the classroom which evaluate content-global and content-specific components of instruction were used simultaneously to predict students' literacy growth over the school year. Additions to these coding systems incorporated observations of classroom materials, specifically the difficulty of reading content, which have been studied independently but never systematically linked to classroom instructional practices or student achievement. Finally, characteristics of students and teachers were related to classroom processes, to evaluate the extent to which these characteristics moderate relations between instruction and reading outcomes

Chapter 2. Review of the Literature

Research has demonstrated in multiple ways that the influence of schooling on students' achievement and socioemotional outcomes provides a critical opportunity for closing gaps in children's academic achievement that begin at an early age (Nye et al., 2004). But to determine what is effective instruction, observations of naturally occurring classroom instruction are needed. Fortunately, a number of attempts have been made in recent years to conceptualize and measure classroom experiences, though few direct comparisons of their predictive power have been attempted. The present study focused on two prominent systems (CLASS and Pathways/ISI) and their independent and interactive impact on children's reading growth.

Overview of Classroom Observation. Classroom observations by teachers, school administrators, and academics for the purposes of research typically fall within a few categories; specifically, these observations may follow the actions of the teacher, the class as a whole, or the interaction between teachers and individual students (dyadic) (Doyle, 1977). Prior to the 1970s, classroom observations focused on classroom features, teacher characteristics, teacher behaviors, and whole class activity (e.g., Ryans, 1961, Ryans, 1963). Brophy and Good (1978) brought about a wave of studies which used a dyadic approach, and pioneered an observation coding scheme which considered the interactions of teachers with individual students. Moving on from this turning point in the field of classroom observation, observation systems have evolved as ways to determine differences between teachers, students, and instruction. Each observational coding system has at its core the researchers' perspective on the question of what constitutes effective instruction.

Current observational coding schemes, which encompass all of these aspects of the teacher, classroom, and students, are still in use in large-scale studies today. For example, the

Early Language and Literacy Classroom Observation (ELLCO) system developed by Dickinson and colleagues (2004) includes measures of classroom structures/furnishings, supplies, and class size. This scale in particular has been widely used in Head Start programs as a measure of classroom structural quality, as well as a method of teacher coaching and professional development. The Classroom Assessment Scoring System (CLASS) developed by Pianta and colleagues (2003, 2007) consists of ratings of teachers' interactions with their students, looking at the average experience of a child in that classroom on a variety of emotional, social, and cognitive dimensions. Dyadic coding schemes, including Brophy and Good's (1996) original system and newer systems such as the Pathways/ISI coding system (Morrison, Cameron, Connor, Strasser, & Griffin, 2005), are widely used to gain knowledge about the unique experience of each child within the classroom. Although these systems focus on the level of individual interactions, rather than an average or overall interaction, the content of instruction considered by dyadic coding schemes still varies significantly.

Although historically both rating systems and field note based systems were widely, and typically separately, used, the majority of current observation systems rely on a combination of these two methods. Systems which use exclusively ratings of observed classroom behaviors and features can be less objective, or provide less information about process, than systems which use a combination of ratings with field notes. Both the ELLCO and CLASS produce ratings of classroom processes, but ask the observer to rely upon field notes, behavioral markers, and relative frequency in order to decide on the correct rating to give. Field note-based systems are commonly used in qualitative classroom research; however, as use of video coding and editing software has become more accessible to both researchers and educators, observational system which use time-based codes and/or counts of behaviors are also on the rise (e.g., Morrison,

Cameron, Connor, Strasser, & Griffin, 2005). These systems can provide nuanced or broad-level information about classroom activities, and provide a metric by which instruction can be measured.

Systematic studies which use multiple observation systems to observe a single set of classroom instruction, such as the MET projects' investigation of classroom observation for the purposes of teacher evaluation and training, are an important step towards creating a comprehensive measure of effective instruction's components (Kane & Staiger, 2012). The MET project, focused on middle school classrooms, has released a number of reports based on this initial study which indicate that the use of multiple measures of the classroom creates the most valid measure of instructional quality. Specifically, this study has used measures of the global classroom climate, teachers' instructional moves and strategies, and content-specific instruction. Linking these measures of the classroom with students' outcomes has provided a wealth of information about effective instruction in the middle school environment.

In addition to the distinction between dyadic and global measures of classroom instruction, teacher and student actions related to outcomes can typically be placed into two broad categories: those which are found in all content-area instruction (content general) and those which are found in only certain types of content-area instruction (content specific). Following this division, coding systems which capture these behaviors can also be categorized as content general or content specific. Below is a review of the most commonly used observation systems from these two categories, and an exploration of how these systems can be expanded and used in concert to explore effective (literacy) instruction in early elementary classrooms.

Content General. Recent work has largely focused on the quality of teacher-student interactions and teachers' instructional moves. At the forefront of this research, Pianta and

colleagues (2004) demonstrated that teacher warmth and sensitivity, classroom organization, and instructional support are all critical aspects of the classroom environment and have been consistently linked to improved student performance in reading ability. Specifically, teachers who provide warm, supportive environments also deliver high quality instruction (e.g., LaParo, Pianta, & Stuhlman, 2004; Mashburn et al., 2006). Examples of warmth include teachers and students demonstrating respectful, close relationships. Supportive environments are both well organized and managed, and also provide high levels of instructional feedback and include higher-order questioning (Pianta et. al, 2008). Moreover, Pianta and colleagues have posited that defining teacher quality through student gains is inappropriate; rather, the provision of emotional and instructional supports should be the metric of teacher quality (Pianta & Hamre, 2009).

Pianta and colleagues have developed and used the Classroom Assessment Scoring System (CLASS) tool not only to evaluate instruction and its effects on student outcomes, but also as a coaching tool. A large-scale study conducted by this research group demonstrated that teachers could be coached, via analysis of videotaped observations of instruction, to improve their scores on the CLASS tool. These improvements were then linked to higher student scores on achievement measures (Pianta et. al, 2008). Indications that changes in teacher performance on this measure are linked to achievement provide strong evidence that the components of the classroom environment that this tool captures are a vital part of effective instruction.

Preliminary results from a recent investigation of first grade instruction found that classroom environment components, specifically warmth/sensitivity and instructional support, significantly positively related to student achievement if students received appropriate reading instruction (Grammer, Guthrie, & Morrison, 2013). First grade classrooms were observed multiple times within the same school year, and coded for both the amount and type of literacy

instruction presented during the school day (Pathways/ISI), but also the classroom environment (CLASS). Analyses of 75 first grade classrooms found that, as expected, literacy instruction predicted student outcomes, particularly the amount of time spent in independent reading activities. Students who took part in the appropriate amount and type of literacy instruction grew more in literacy skills over the school year, and they also had an extra boost in reading achievement outcomes if their classroom was rated as highly warm and supportive. If students did not receive the appropriate amount and type of reading instruction, they experienced decreased growth compared to their peers. However, this decreased growth was mitigated in classrooms which had higher levels of instructional support. These teachers may have posed higher order thinking questions to students, used advanced language, or engaged students in feedback loops regarding instructional foci. These supports may have made instruction more challenging for advanced readers, or may have provided extension activities for those students.

Thus, it may be that a positive, supportive environment is not sufficient for student achievement in the absence of the appropriate amount and type of instruction, but does support correctly targeted instruction. The ability of a teacher to provide individualized instruction is an important component of teaching, and is highly linked to student outcomes. Continuing to compare both the amount and type of instruction provided, in addition to the environmental supports available, is important as researchers continue to determine the characteristics of effective literacy instruction. The present study used this observation protocol in enrolled classrooms in order measure the classroom climate and processes that all students experienced, and linked these observations to student performance.

Content Specific - Literacy. Moving on from these broad conceptualizations of the classroom environment, researchers have begun to examine how specific instructional activities

and actions are related to student reading ability. Indeed, numerous studies have found that there are distinct amounts of instructional activities that are most appropriate for struggling and advanced readers (e.g., Connor, Morrison & Katch, 2004; Sonnenschien, Stapleton, & Benson, 2010). Effective teachers use their knowledge of students' abilities in an effort to build students' reading skills (Piasta, Connor, Fishman, & Morrison, 2009; Sameroff & MacKenzie, 2003). Thus, it is unsurprising that children's initial knowledge is an important consideration in instructional effectiveness (e.g., Connor, Morrison, & Katch, 2004; Connor, Morrison, & Slominski, 2006; Morrison, Bachman, & Connor, 2005). Teachers must take into account the skills children bring into their classroom in order to create the largest gains in student achievement over the school year.

In order to classify effective instruction, Morrison, Connor and colleagues have found that there are distinct amounts of instructional activities that are appropriate for struggling and advanced readers (e.g., Connor et. al, 2009). These researchers have identified the following two categories: type of instruction (teacher or child is directing the activity) and focus of instruction. The latter reflects whether instruction proceeds at the level of letters and sounds (code-focused) or at the level of words and sentences (meaning-focused). Morrison, Connor and colleagues have repeatedly demonstrated that students with below average literacy skills benefit most from teacher-managed code-focused activities (e.g., phonics), whereas students with above average literacy skills benefit most from meaning-focused activities (e.g., independent reading) (Connor et. al, 2004; Connor et. al, 2006; Morrison et. al, 2005).

More recent work by this research team has found that the effect of management of meaning-focused activities for students with above average reading skills differs based on students' grade-level (Guthrie & Morrison, 2012). Younger highly skilled readers (e.g.,

preschoolers and kindergarteners) benefit most from teacher-managed meaning-focused instruction. Older highly skilled readers (e.g., first grade and above) benefit most from child-managed meaning-focused instruction. These results are supported by previous findings that preschool and kindergarten students benefit largely from teacher-managed instruction because they do not have the necessary self-regulation skills to effectively learn independently (Ponitz & Rimm-Kaufman, 2011). Students in first grade and beyond, who on average have higher self-regulation skills than preschoolers, are able to focus their attention during independent activities, and thus can benefit from child-managed instruction.

Additional studies have reinforced these findings in nationally representative samples (e.g., Sonnenschien, Stapleton, & Benson, 2010). Sonnenschien and colleagues (2010) used the Pathways/ISI conceptualization of code- and meaning-focused instruction to explore the connections between teachers' reports of time spent in these types of literacy activities and student achievement outcomes in the Early Childhood Longitudinal Study dataset. The researchers modeled student growth in reading ability from kindergarten through fifth grade, using teacher reports of instructional amount and type to predict student growth. Results supported the general child-by-instruction interaction findings from Morrison, Connor and colleagues (e.g., Connor, Morrison & Katch, 2004) in showing that students with lower reading skills benefit from more time spent in code-focused instruction, whereas students higher reading skills benefit from more time spent in meaning-focused instruction.

An intervention study demonstrated improved student reading skill when teachers were provided with supports and knowledge regarding how to individualize instruction using these two dimensions (Connor et. al, 2009). Specifically, teachers were provided with software which uses student reading achievement scores to predict the amount and type of instruction that will

improve student reading achievement scores by one grade level. Teachers who most closely followed the software's recommendations, providing more teacher-managed code-focused instruction to struggling readers and more child-managed meaning-focused instruction to advanced readers, had large gains in student reading skills (Connor et. al, 2009). Further studies using this intervention software have replicated these results from kindergarten through second grade (e.g, Connor & Al Otaiba, 2012).

Expanding the focus of Pathways/ISI: The role of content. Recent research involving classroom observations in kindergarten found that more instructional time spent in teacher-managed code-focused instruction predicted significantly lower growth in literacy achievement test scores for above average readers (Guthrie & Morrison, 2012). Connor, Morrison and colleagues (2004; 2006) also found that reading scores of above average readers plateaued when they received more teacher-managed code-focused instruction.

These findings that particular types of instruction are not just ineffective for some students, but that they limit literacy knowledge growth, are puzzling. More detailed research may indicate what aspects of code-focused instruction are least effective; for example, teachers who engage in these activities may have classes made up of largely below grade level readers. Alternatively, Cunningham, Zibulsky, Stanovich, and Stanovich (2009) observed that teachers who were less knowledgeable about phonics preferred to allocate twice as much time to literature (or meaning-focused) activities than teachers who were more knowledgeable about phonics. In contrast, teachers who were more familiar with phonics preferred to spend three times as long on instruction that was focused on letters and sounds versus literature activities. Based on these findings, further exploration of the reason behind the ineffectiveness of these activities for high ability readers is needed, as well as an exploration of how this kind of instructional time could be

made more relevant to highly skilled readers. One potential area of study which may be related to these findings is the content of these activities (e.g., lesson objectives, text difficulty).

Literacy Instruction Content

The worksheets, storybooks, exercises, and examples that teachers use during literacy instruction comprise an important part of the instructional experience. Teachers use a wide variety of sources, including trade books, websites, fellow teachers, and previous experience, to obtain and alter materials to suit their needs. Researchers know relatively little about the processes which link use of these materials to teachers' goals for individual students (Piasta et. al, 2009). Research by Dickinson and colleagues, found strong relations between the quantity, availability, and quality of books in preschool and prekindergarten classrooms with student emergent reading skills (Dickinson et. al, 2001). Findings from this study imply that students who are encouraged and able to spend more time with high quality texts during school have better reading skills than their peers in classrooms with fewer, or less diverse, books to choose from. Aside from these studies which use the number and broadly defined quality of books, researchers have yet to investigate specific links between the content of instruction (that is, materials such as texts and worksheets) and student achievement outcomes.

Previous research has identified the ways in which grade- and age-leveled readers, books, and supportive texts might contribute to literacy skill development (e.g., Hiebert, Martin, & Menon, 2005). Teachers may be using these leveled materials in order to more exactly prescribe instruction to meet individual students' needs. However, to date research into texts in elementary school classrooms has focused on the particular aspects of text that make it more or less accessible to beginning readers (Hiebert et. al, 2005), as well as the amount of time students spend interacting with text in general (Brenner & Hiebert, 2010).

Hiebert and colleagues looked at the number of unique words on a page and within a text, as well the frequency of those words and the structure of sentences within the text. They found that the difficulty of text is strongly linked to these three dimensions, and that these three dimensions are typically related to publisher-determined grade- and age-level suggestions. Moreover, the components of text that make it more or less difficult (e.g., unique words, word frequency, sentence structure) are strongly related to word learning and reading skill development (Cunningham et. al, 2005; Hiebert et. al, 2005). Often these aspects of text are considered a measure of readability. Other established systems, such as Flesch-Kincaid, Fog Index, Coh-Metrix, and Lexiles, have used similar components of text in order to create difficulty ratings or scores. Hiebert and Pearson (2010) compared these current methods through their TeXT project. Generally, Flesch-Kincaid and Fog both use word difficulty, words per sentence, and text length in the creation of a score (Flesch, 2006). Similar aspects of text are used in Coh-Metrix (McNamara, Graesser, Cai, Kulikowich, & McCarthy, 2010); however, this system also uses measures of cohesion, type of text (e.g., narrative), and syntactic complexity. Lexiles (Stenner, Burdick, Sanford, & Burdick, 2007) are largely computed based on average word difficulty and words per sentence (semantic and syntactic difficulty, respectively).

In reviewing these measures, Hiebert and Pearson used each scale to measure the readability/difficulty of the same set of texts. This systematic comparison of the scales found that the all four of the systems consistently rated texts such as *Cat in the Hat* and *Caldecott Medal* picture books (which Hiebert terms “trade texts”) as the most difficulty, while rating books such as *Dick and Jane* (which Hiebert terms “historical texts”) as the easiest. The ratings of specific books were not necessarily the same across all measures; however, the general grouping of texts by degree of difficulty was repeated by each system. This review concluded that the information

provided by these rating systems is general in nature, but may not necessarily be useful in suggesting the ideal texts for beginning readers based on their knowledge/skills.

Hiebert and Pearson (2010) indicate that an important next step is addressing the link between students and texts, which has not been addressed in the literature. Such a combination of student experience with texts would enable the creation of a readability measurement system that could be used to identify texts that might help readers master particular skills. In addition, less is known about the exercises and materials teachers use to teach letter- and word-level skills, and there is a lack of research linking text difficulty with student outcomes. Teachers make use of a variety of letter-based activities to instruct students in their knowledge of phonics and decoding (Piasta et. al, 2009). Dickinson and colleagues indicated that preschool and prekindergarten teachers who use the appropriate terminology for phonemic awareness activities, and who incorporate these activities into a classroom schedule along with meaning-focused activities, have students with higher emergent literacy skills (Dickinson et. al, 2001). Few if any studies have specifically examined these activities for difficulty and content. Findings that these code-focused activities are ineffective for advanced readers may imply that teachers typically use simple, easy exercises and materials when presenting code-focused instruction. Because advanced readers do not necessarily have complete knowledge of phonics and decoding, perhaps more difficult code-focused activities might be more appropriate for those students.

Incorporating this information into the existing Pathways/ISI coding system would provide valuable information about the processes underlying the broad activity categories previous studies have established as the most effective for particular students. To this end, the present study collected information on the books and texts used during instruction, as well as on worksheets and teacher-created writing, for inclusion in analyses linking coded literacy

instruction to student reading outcomes. Specifically, the difficulty of books and texts was measured using publisher-determined age- and grade-level guidelines. Worksheets, and in addition board examples, were coded using the dimensions of print established by Hiebert and colleagues; specifically, the number of unique words, the frequency of words, and the structure of the assignment will all be considered in determining the difficulty of content presented through worksheets and teacher-created writing.

Teacher Characteristics

Various aspects of teachers and their classes have also been linked to student achievement and effective instruction. A major component of effective instruction is teachers' ability to determine students' reading skill strengths and weaknesses (Piasta, Connor, Fishman, & Morrison, 2009; Sameroff & MacKenzie, 2003); thus the accuracy of teachers' ratings of student ability may be related to their instructional practices. Moreover, teacher experience has been linked to student outcomes (Huang & Moon, 2009), and these experiences likely play a role in the links between instruction and student outcomes.

Teacher ratings of student ability. In order to individualize instruction, teachers must use formal and informal assessments in order to gauge the abilities and needs of their students. A number of prior studies of teacher knowledge and instruction have found that teacher assessments and ratings of students are often problematic, leading to incorrect judgments of what amount and type of instruction to provide. Begeny, Krouse, Brown and Mann (2011) found that teachers only accurately identified just over half of students' reading levels. Hecht and Greenfield (2002) found that some factors that do not influence a child's reading attainment (e.g., classroom behavior) can sometimes influence teacher's predictions of reading ability, leading teachers to misclassify students based on erroneous information. Shepherd (2011) found in a

similar study that teachers hold lower expectations of boys as compared to girls with regard to reading ability. These misclassifications, which appear to be relatively common, could hinder teachers' ability to provide appropriate instruction to students. Indeed, a teacher who believes a student with high decoding skills needs additional work on letter sounds would not be able to effectively increase student reading achievement. Moreover, Pianta and colleagues (2002) found that teacher ratings of student ability were strongly linked to classroom climate and teacher support of students during instructional activities.

This body of research indicates both that software such as that used in Connor and colleagues' intervention is vital to aiding teachers in improving instruction, and also that examining teacher ratings of student reading abilities, in addition to student achievement test scores, may explain the type and amount of instruction students receive and its effects. The use of the software eliminates teacher bias or misjudgment, as it uses objective test scores in order to prescribe instruction. Because teachers observed during naturally occurring instruction will be using their own judgments in order to align instruction with student needs, the current study asked teachers to rate student skills over the school year on multiple dimensions of reading ability and learning-related skills. The reading ability dimensions included oral language skills, listening comprehension, letter knowledge, phonemic awareness, and reading comprehension, and are strongly linked to the independently tested student outcomes. In addition, teachers will also be asked to rate students' self-regulation and motivation for reading. These ratings will be compared with initial student test scores, and will also be linked with instruction and student outcomes.

Teacher experience. Studies of teacher education programs have found consistently varied results regarding the importance of teacher education to effective instruction. Indeed,

Hanushek and colleagues (1992) found that teachers vary greatly in their ability to improve student achievement, and that this variance is not predicted by easily-measured teacher characteristics such as highest degree of education or certification status. Huang and Moon (2009) note that U.S. policy has emphasized the importance of highly qualified teachers in the classroom, without adequately defining exactly what attributes are indicative of being highly qualified. Moreover, the characteristics traditionally associated with this standard – education, knowledge – are not linked conclusively to children’s outcomes (e.g., Darling-Hammond, 2000; Croninger et. al, 2007).

One characteristic that Huang and Moon (2009) have linked with student achievement is teachers’ years of experience teaching at grade level. The authors posit that experience teaching the same material as well as similarly-skilled students may enable teachers to become more adept at identifying student strengths and weaknesses. Moreover, teachers who have taught the same grade for a number of years may have developed a more advanced, nuanced understanding of the material and may be better equipped to push students to higher levels of achievement. Other prior studies examining teacher effects across multiple years of schooling found that teachers with more experience were linked with improved reading comprehension skills from preschool through second grade (Guthrie, Connor, & Morrison, 2011). Therefore, it seems likely that teacher experience in particular may be an important component of effective instruction.

Improving Individualized Instruction

From this collection of research, it is clear that there are ways in which recommendations for individualizing instruction can be improved. Assessments of the effectiveness of classroom instruction which ignore dimensions of warmth and support may be missing out on important aspects of the classroom which can contribute to the positive effects of properly individualized

instruction. Moreover, the observations and prescribed teaching recommendations made using Connor, Morrison and colleagues' system do not take into account the difficulty of the content found in the materials and examples teachers use to support their instruction. These subtle differences may influence child by instruction interactions, as this system would classify students engaging in similar activities which are in fact not comparable in terms of content and text difficulty. Therefore, this study hopes to explore this other facet of instructional activities.

In addition, the current study explored the accuracy of teacher ratings of students' ability, and the degree to which teachers use their own ratings to individualize instruction. This aspect of teacher knowledge is an important component of effectively prescribed instruction, and further research in this area may help researchers to develop training programs to improve the accuracy of teacher assessments of student ability. Finally, this study will investigate the role of teacher experience in teachers' ability to individualize instruction effectively. Thus, this study has been designed to examine additional aspects of instruction and the classroom that may be related to effective reading instruction and to student outcomes.

Hypotheses

1. Teachers will individualize instruction effectively along dimensions of instruction (e.g., focus, management) that have been linked to student achievement growth.
2. In linking measures of classroom climate, reading instruction, and student outcomes:
 - a. There will be limited links between measures of reading instruction and measures of classroom climate.
 - b. Classroom climate will be uniquely linked with student achievement growth; specifically, emotional support and instructional support will predict students' gains in reading achievement.

3. The level of content difficulty, specifically the difficulty of books, worksheets, and teacher-created print, will be related to reading instruction and student achievement growth. Students will be interacting with books at a similar difficulty rating to their reading skills.
4. In linking teachers' ratings of students' skills with achievement and instruction:
 - a. Teacher ratings of students' skills will be moderately positively related to students' standardized achievement test scores.
 - b. Teacher ratings of students' skills will be moderately related to literacy instruction and classroom environment.
5. More experienced teachers will provide more literacy instruction and have warmer, more supportive classrooms.

Chapter 3. Method

Participants

The participants in this study were seventeen first grade teachers and their students; teachers were recruited through their principals and contact with district officials and are nested into six schools. All schools taking part in the study were public elementary schools from around southeastern Michigan. The schools provided information regarding the percentage of students receiving free and reduced price lunch during the school year in which data was collected; this percentage ranged from 23.80% to 73.10%. Teachers had on average 15.54 years of experience teaching (range from 4 to 38 years), and 8.39 years of experience teaching first grade (range from 1.5 to 20 years). Fifteen of the teachers had a Master's degree in education, and all held Bachelor's degrees in areas related to education, reading/language arts, or mathematics.

Two hundred and thirty-six students enrolled in the study, and 233 students ultimately participated fully in the study and had usable data (107 female). Two students moved away after the first time point and did not participate in spring observations or testing, and one student who had been retained in first grade twice was an outlier with regard to reading achievement test scores. The average student age at the beginning of the study was 6.55 years ($SD = .439$). Of these students, 77.6% were White, 9.5% were Black, 7% were Asian, 4.3% were Latino/a, and 1.6% were Native American.

During the course of the study, one teacher requested that her classroom not be observed, but did allow her students to be assessed in the fall and spring of the year. In addition, one classroom had 50 enrolled students as well as two head teachers and one student teacher. Both head teachers' background information was used; however, the student teachers' data were excluded because a different student teacher was present at each observation.

Procedure

Classroom teachers distributed consent forms during the first week of school. Within the next six weeks students with parental consent were assessed on three reading achievement measures and two measures of executive function by trained undergraduate and graduate research assistants. Simultaneously, teachers rated their participating students' skills on a variety of reading and reading-related skills.

During the second (fall) and third (spring) marking periods research assistants conducted classroom observations using the Classroom Assessment Scoring System and an adapted version of the Pathways/ISI Classroom Coding System. Teachers provided any materials used during the lesson in order to supplement the observations, including copies of handouts and worksheets as well as the titles of any books or materials used during the lesson.

During the fourth marking period students were assessed again on the same measures of reading achievement and self-regulation. Materials and content collected during observations were coded by trained research assistants using internally developed coding schemes. Finally, in order to gather information about students and teachers, questionnaires were sent out to gather background information.

At the conclusion of each assessment time point, all student testing protocols were double scored and double entered by the research team. In addition, all observation narratives and codes were double entered by the research team, and all variables created for use in analyses and models were z-scored in order to standardize across variables from a variety of sources.

Materials

Student Assessments

Reading ability assessments. Children took part in one-on-one assessments with researchers in the fall and spring of each study year. These assessments consisted largely of achievement measures from the Woodcock-Johnson Tests of Achievement III (WJ-III). These tests assess a broad range of cognitive, language, academic and social skills. This study focuses on students' achievement scores on the Letter-Word Identification, Picture Vocabulary, and Passage Comprehension scales from this larger battery of tests. Mather and Woodcock (2001) describe each of these scales in terms of the items, difficulty, and reliability amongst age groups of interest. The Letter-Word Identification (LW) scale measures children's ability to identify increasingly difficult printed upper and lower case letters as well as words. It tests decoding and pronunciation specifically, and has a reliability of 0.91 in elementary school samples. In the sample of students assessed for this study, Chronbach's alpha was 0.81 for this scale.

The Picture Vocabulary scale (PV) asks children to identify the word represented by a picture, measuring oral language development and lexical knowledge. The difficulty of the words increases through the scale, and this scale has a reliability of 0.77 in elementary school samples. In the sample of students assessed for this study, Chronbach's alpha was 0.72 for this scale.

The Passage Comprehension scale (PC) uses a cloze procedure and requires students to identify the meaning of a picture, phrase, or paragraph. The difficulty of the scale increases with each item as the vocabulary, syntax, and semantic clues become more and more complex. The reliability of this scale is .83 in elementary school samples. In the sample of students assessed for this study, Chronbach's alpha was 0.78 for this scale.

Two different forms of each of these tests are available, and students received a different form in the fall and spring. Student assignment to test form was counterbalanced, so that half of the sample received each form in the fall and spring.

Training and Fidelity. In order to train undergraduate research assistants to use these tools a presentation and demonstration of the testing materials was given by the graduate student leading the project prior to data collection beginning. The research assistants were required to practice giving and scoring the achievement tests over the course of one week. A fidelity of implementation checklist and a mock student testing protocol were created, and each research assistant was assessed for fidelity and reliability of test administration. A score of at least 90% on the checklist was required prior to administering these tests in the field. A similar fidelity of implementation drift test was completed prior to the second assessment time point.

Classroom Observations

Classroom environment. For the purposes of this study, the Classroom Assessment Scoring System (CLASS) was used in order to capture aspects of the classroom not related directly to instruction. Pianta, LaParo, and Hamre (2008) presented the CLASS Framework as a method for characterizing the structure and nature of teacher-child interactions. These aspects of interactions were posited to contribute positively to students' development. The CLASS system provides a characterization of three broad domains (**Emotional Support, Classroom Organization, and Instructional Support**) and 10 specific dimensions (Positive Climate, Negative Climate, Teacher Sensitivity, Regard for Student Perspectives, Behavior Management, Productivity, Instructional Learning Formats, Concept Development, Quality of Feedback, and Language Modeling), each of which are presumed to be important to students' academic and/or social development.

Developed based on in part on literature on effective teaching practices as well as extensive classroom observations conducted in large-scale observational studies of early school settings (e.g., NICHD ECCRN, 2002), the CLASS was created to provide a standardized measurement tool and language for describing classroom quality in pre-K and early elementary school classrooms. Using the CLASS, interactions between teachers and children are characterized by observers. Evidence based on the use of this measure in a broad array of research inquiries suggests that CLASS captures aspects of teacher-child interaction that are stable across a specific day, across days, across students, and across content area of instruction, thus providing a reasonable estimate of features of a teachers' behavior that appear stably characteristic of his or her interactions with students (Pianta, LaParo, & Hamre, 2008). Thus, the CLASS provides a content-general perspective on classroom processes.

CLASS Domains and Coding Dimensions. The domain of ***Emotional Support*** was created to capture the extent to which teachers provide support for the social and emotional development of their students, and provide an environment which encourages students to relax and enjoy learning. This domain is comprised of four dimensions including: Positive Climate, Negative Climate, Teacher Sensitivity, and Regard for Student Perspectives. Together, these dimensions reflect the emotional connection between the teacher and student, shared enjoyment observed in the classroom, teachers' responsiveness to the concerns of students, and degree to which teachers incorporate students' interests and points of view in daily activities.

Classroom Organization dimensions capture the processes in the classroom that are related to the way in which students' behavior, time, and attention in the classroom is managed. Support for the importance of classroom organization is derived, in part, from research on children's development of self-regulatory skills (Blair, 2003; Raver, 2004), and literature

demonstrating the important role of the classroom setting for the development and expression of children's regulatory skills. Classroom Organization consists of three dimensions including Behavior Management, Productivity, and Instructional Learning Formats. The dimensions in this domain reflect teachers' effectiveness in monitoring and redirecting behavior, routines and organization of activities within the classroom, and the extent to which teachers engage children in and facilitate lessons.

Focusing on aspects of the classroom environment that foster children's cognitive development, the *Instructional Support* domain consists of dimensions which measure teachers' support for the development of students' usable knowledge. The dimensions within this domain, including Concept Development, Quality of Feedback, and Language Modeling, do not capture the curriculum that teachers use or the specific activities in which the students are engaged. Instead, each of these dimensions are coded based on the specific interactions in which teachers promote higher-order thinking skills, extend students' understanding through their feedback, and stimulate language development through conversation and use of advanced language.

Coding with the CLASS. Each dimension is composed of several behavioral markers which observers are trained to identify in the classroom; the presence, absence, and quality of these markers are operationalized by observers with a score on a 1-7 scale. To assess the classroom environment using the CLASS, a minimum of two hours of instruction is observed in 30-minute cycles. To obtain ratings of each dimension, coders observe instruction in 20-minute intervals while taking notes about what is occurring in the classroom. At the end of the 20 minutes observation cycle, coders then assign scores for each dimension during the 10 minute coding cycle. These scores range from 1 (minimally characteristic) to 7 (highly characteristic) capturing the quality of each dimension's behavioral markers as observed during the 20 minute

observation cycle. Coding for each block is to be considered separately across the total time observed in any individual classroom, and ratings are based on the trained coders' judgments regarding the exchanges and activities seen during the observed period.

Training and Reliability. Reliability for this measure was achieved in two separate ways. First, the graduate research assistant conducting the CLASS observations had been trained and certified through TeachStone, the company which owns and distributes the CLASS tool. Individuals using the CLASS must complete extensive training in the observational tool and reach a level of reliability coding a set of master-coded videos. Later drift testing through TeachStone confirmed that the research assistant was reliable (agreement over 80% of scores across five cycles, and no more than one disagreement per dimension across five cycles) one and two years after this training, which is the only requirement TeachStone has in place to maintain the reliability of certified observers over time. In addition, because this research assistant observed classrooms using the CLASS tool and the adapted Pathways/ISI coding system, further reliability testing was conducted to ensure that coding with these two tools simultaneously was not causing biased measures of the classroom. This testing consisted of recoding with both tools 15 videotapes of classroom observations which had previously been coded by master/certified coders with each tool separately. The results of this recoding demonstrated 92% agreement with master Pathways/ISI codes and 94% agreement with certified CLASS codes.

Typically, the three larger dimension scores created from the 10 domain scores are used in studies which employ the CLASS coding scheme. Correlations amongst the domains and dimensions are shown in Table 8.

Pathways/ISI Classroom coding. Trained researchers also live-coded classroom instruction using an adapted version of the Pathways/ISI system (Morrison, Cameron, Connor,

Strasser, & Griffin, 2005). Traditional observational coding using Pathways/ISI employs Noldus Observer Pro software, videotapes of the classroom from multiple angles, and well-detailed time-stamped field notes. The system captures 4 dimensions of all activities: (1) *management* of students' attention (teacher- or child-); (2) *grouping* of the activity, such as whole class, small group, and independent; (3) *focus* of the activity (code- or meaning-); (4) *duration* of activity. These dimensions occur simultaneously, so that any activity with a duration of 15 seconds or longer is defined along all four dimensions. Previous investigation using the system on the Noldus software have achieved high inter-rater reliability ($\kappa = .78-.92$) (Landis & Koch, 1977).

Based in part on literature which linked student outcomes to time on task as well as student engagement, (Arlin, 1979; Karweit & Slavin, 1981; Pressley et al., 2001; Meece, Blumenfeld et al., 1988; Ryan & Patrick, 2001; Skinner, Wellborn et al., 1990), the *management* dimension allows for the classification of who is focusing the child's attention on the learning opportunity. This coding system allows for several descriptors of the management of activities, two of which are of importance in this study: child-managed (CM, child independent engages in activities without the support of a teacher), and teacher managed (TM, teacher is involved in directing students' attention).

The second coding dimension, *grouping*, is used to document whether students and the teacher are working together as an entire class (whole class), in smaller groups, or independently. The third dimension of the coding scheme, *focus*, captures the learning objective of an activity. In the larger coding system, *focus* is comprised of hundreds of specific activity codes which describe nuanced aspects of lessons; for the purposes of this study, larger categories describing these activities will be used: code-focused (CF, activity proceeds at the letter or individual word

level with focus on sounds, syllables, etc.), and meaning-focused (MF, activity proceeds at the word, sentence, or story level with focus on definitions, comprehension, composition, etc.). The fourth dimension, *duration*, is captured through the timing of codes and reflects the length of time during which students are engaged in activities.

The coding scheme allows all four dimensions (management, grouping, focus, and duration) to be coded continuously throughout the observation for each individual child within a classroom. Individual activities often contain multiple codes; for example, a teacher may work one-on-one with a student listening to the student read aloud a text, which in the original coding scheme would be coded as a teacher-managed, individual, student read-aloud activity with the duration indicated by the timing of the code. The student might read uninterrupted until the teacher asks a comprehension question, points out initial consonant sounds, or asks the child to interpret the picture on the page. Each of these smaller activities within the larger read-aloud task would receive separate codes so long as they lasted at least 15 seconds each. This example illustrates the continuous nature of the coding scheme, and the detail with which various aspects of individual activities are captured within the Pathways/ISI coding system.

Continuing the above example of a student reading aloud to a teacher, previous studies employing the Pathways/ISI system would have collapsed such a code along with many others into one variable for the purposes of analyses (specifically, into a teacher-managed meaning-focused). Because the level of detail provided by this coding process is not utilized fully in these studies (indeed, the grouping of the activity is entirely ignored in this kind of analysis), and the time involved in capturing that amount of detail is extensive, an adapted version of the Pathways/ISI coding system was created. This adapted coding system uses three of the four original dimensions (**management**, **focus**, and **duration**) and includes a narrative which

captures the **grouping** of the activity as well. Coders used pen-and-paper as well as stopwatches to record a description of activities as they were observed during live classroom observations. Start/Stop time as well as **duration** were recorded for all activities lasting longer than 15 seconds, and written notes describing the activity were written in a running narrative. Coders also determined the **management** and **focus** of the activity. Because multiple coders were present in each classroom during an observation, there was opportunity for group discussion and consensus regarding codes and timing, which improved accuracy.

Some aspects of the Pathways/ISI coding system prove to be barriers to implementation; specifically, the use of video and audio equipment in the classroom, as well as the time intensive coding process. Many teachers and parents have become wary of permitting researchers to videotape their children, stemming largely from professional and personal concerns for privacy. Recruiting schools for studies which involve video and audiotaping of lessons can be difficult. Because the Pathways/ISI coding system requires a lengthy observation, the number of cameras and digital data storage required can also be a limiting factor. In addition to the difficulties that can be encountered while trying to collect data, processing and using the collected video data via the Pathways/ISI coding interface in Noldus Observer is a lengthy and involved process. Coding a single two-hour videotape can require four to six hours of coding, and thus the amount of time needed to obtain usable coded data can be prohibitive due to time and staff costs.

As a result of these two limitations in particular, an adapted coding system was created as part of the larger study described here. The Pathways/ISI coding system provides great detail about the instruction observed; indeed, reading instruction coded in this way is classified into 26 separate kinds of activities (e.g., initial consonant, handwriting practice). When this coding system is used to analyze data from both studies of naturally occurring instruction and

intervention work, activities are summed together into four categories (teacher- or child-managed/code- or meaning-focused) for the purposes of analyses and description of instructional effectiveness. The adapted coding system uses only these four category codes to describe classroom instruction, allowing coding to be done live in classrooms during typical instruction. Additional narratives and notes of student groupings, activity components, and teacher/student writing will also comprise an important aspect of this adapted coding system.

One advantage of the original coding system, which posed a major limitation for the current study, is the ability to code multiple students from a single classroom. In order to conduct live coding of a child's behavior, one observer must be present for each child. This drastically limited the number of children that could be observed at one time. Based on prior work which has used the Pathways/ISI system to produce recommendations of amount and type of instruction each child should receive based on incoming reading achievement scores, it was decided that three children would be observed in each classroom. All participating students were classified as above, at, and below grade level in terms of their reading skills. These classifications were based on fall reading achievement scores in decoding, passage comprehension, and vocabulary. Percentile ranks, which compare a child's score on the achievement test to a nationally normed sample of students of the same age and grade, were computed for each score for each student. Students were considered to be above grade level if their scores on all three tests fell above the 70th percentile for their age and grade. Similarly, students were considered to be below grade level if their scores on all three tests fell below the 40th percentile for their age and grade. This meant that students with scores on all three tests between the 40th and the 70th percentiles were considered to be at grade level. One child from each classification was observed during each observation, with the same child being observed in both fall and spring.

Training and Reliability. Nine classroom observers were trained on the adapted measure. This training involved coding nine videotapes of classrooms from prior studies (five videotapes prior to the fall classroom observation time point, and four videotapes prior to the spring classroom observation time point). After viewing and coding the videotapes, codes were compared to a set of master codes. An inter-rater reliability of at least 85% was achieved between all nine observers with the master codes on all nine videotapes.

Content difficulty coding. The narrative captured during classroom observations included many details about the activities students engaged in; in particular, the title of books and the use of worksheets and teacher-created writing were recorded in these narratives for further exploration and coding.

Worksheet coding. Worksheets (and teacher-created writing, which was not seen in the classrooms during observations) were coded using a system which captured three components: number of directions, number of words/number of unique words, and word frequency or difficulty using the Educator's Word Frequency Guide (NICHD & Zeno, 1995). These components measure the amount of material children need to hold in mind while completing an activity, as well as the difficulty or complexity of the material presented. Two undergraduate research assistants and a former teacher assisted in the development and implementation of this coding system. The former teacher was considered the master coder, and the two undergraduate research assistants reached 91% reliability with her across all three components of the coding system on a set of 15 sample worksheets drawn from available first grade curricula materials. The two undergraduate coders both coded all collected worksheet materials, and disagreements were settled through discussion and mutual decision making.

Book/text difficulty coding. Researchers and teachers collected the titles of books students read or interacted with during classroom observations in the running observation narrative. A number of rating systems are available for coding the difficulty of books and texts, including Flesch-Kincaid, Fog Index, Coh-Metrix, and Lexiles. These rating systems use similar aspects of text as those incorporated into the researcher-developed worksheet coding scheme (e.g., average sentence length, number of unique words, average word difficulty). Using a subset of 50 of the observed book titles, research assistants searched publicly available information from publisher and distributor websites in order to obtain all available ratings for each title. In some cases (approximately 5 of the 50 books) research assistants called or emailed publishers with requests for available ratings. The results of this search indicated that Flesch-Kincaid and Fog Index scores were available for 46 of the 50 books within the subset, Coh-Metrix scores were available for 6 of the 50 books within the subset, and Lexile scores were available for 39 of the 50 books within the subset. In addition, calculators which could provide Flesch-Kincaid scores based on entered text were also available to the research team. This was an important resource, as many of the books observed in classrooms were no longer in print, or were curriculum-specific texts which the research team photocopied at the conclusion of observations. Given previous findings regarding the similarity of results when using these scales on a large number of texts, a decision was made to use Flesch-Kincaid scores as the rating of text difficulty for this study.

The Flesch-Kincaid measure of text readability and difficulty was originally developed in the 1970s by Naval researchers in order to determine the difficulty of technical manuals (Flesch, 2006). It takes into account the average number of words per sentence as well as the average number of syllables per word within a particular text. It is currently used by publishers in order

to target a particular readability level, and many researchers use it in order to make sure consent and survey materials will be easily understood by their study participants. Scores align with grade levels, with a lower score indicating an easier to read text.

Parent and Teacher Questionnaires

Student ratings. Throughout the year, teachers were asked to rate the skills of their students on seven dimensions of reading and learning skills. Teachers were asked to rate students on a scale from 1-5, with 1 representing a student in the bottom 20% of all first grade students, and a 5 representing a student in the top 20% of all first grade students, with regard to the skill in question. The dimensions include oral language skills (ability to communicate ideas), listening comprehension (how well a child can understand stories read aloud), letter knowledge (knowledge of letter names and printing), phonemic awareness (knowledge of sounds of letters, digraphs, etc.), reading comprehension (ability to understand stories and make predictions/connections to text), self-regulation (ability to focus attention, follow directions, and avoid misbehavior and distractions), and motivation for reading (interest in reading and learning to read).

Teacher background questionnaire. In the spring, researchers asked teachers to provide basic information about their classrooms (e.g., number of students, number of aides, curriculum used), as well as information about the teacher's own experiences (e.g., education, professional development, years of teaching). The response rate to these questionnaires was 100%.

Parent and child background questionnaire. Parents and guardians were asked to respond to questions about caregiver education and employment, ethnicity, family size, and home language and reading environments. The response rate was 78%, with questionnaires returned entirely via mail.

Chapter 4. Analyses and Results

Description of Student Achievement

Student Decoding Skills. Students entered the first grade school year with an average grade equivalent score on decoding of 1.85 (Range=0.40 to 6.70). Their average percentile rank was 69.92% (Range=4.00% to 99.99%). Students showed growth in this measure over the school year, and at the spring time point the average grade equivalent score for decoding was 2.59 (Range=0.70 to 9.80). Their average percentile rank was 75.37% (Range=3.00% to 99.99%).

Student Passage Comprehension Skills. At the beginning of the school year, first graders enrolled in this study had an average comprehension grade equivalent score of 1.32 (Range=0.00-3.80). The average percentile rank for these students on this was 53.25% (Range=0.10%-99.70%). At the end of the school year, students' average grade equivalent score had grown to 2.01 (Range=0.10 to 6.10). Their average percentile rank was 61.59% (Range=0.20% to 99.99%).

Student Vocabulary Skills. On the Picture Vocabulary measure, students in this study began first grade with an average grade equivalent score of 1.97 (Range=0.00-6.60). The average percentile rank for students at this time point was 61.78% (Range=1.00%-99.00%). Scores increased by the spring time point, with students earning an average grade equivalent score of 2.42 (Range=0.10 to 8.10). Their average percentile rank was 62.94% (Range=6.00% to 98.00%).

Research Question 1: Individualization of Instruction

In order to investigate whether or not teachers were providing individualized instruction for students, a formula created by Connor, Al Otaiba, and colleagues (2011) was used to determine the amount of each type of instruction (identified via the Pathways/ISI system) each student would be predicted to need at the fall and spring time points for optimal growth. This

formula was conducted to determine the appropriate amount of Teacher-Managed Code-Focused (TMCF) instruction for each student in October and April of the first grade year:

$$\text{TMCFRecc}(\text{in minutes}) = (\text{Target Outcome (TO)} - (0.2 * \text{Letter-Word Reading Grade Equivalent (LWRGE)}) / (0.05 + (0.05 * \text{LWRGE}))) / 2.2 + 2.$$

Where LWRGE, decoding grade equivalent score, is used to determine the number of minutes of instruction a child should receive, and TO = fall LWRGE + 0.9 with a minimum GE of 2.0. The recommended amount goes down 30 seconds per month. The function is non-linear with exponentially more TMCF instruction recommended for students whose fall scores fall below a GE of 0.5.

The following formula was then used to determine the recommended amount of Child-Managed Meaning-Focused (CMMF) instruction for each student in October and April of the first grade year:

$$\text{CMMFAmount}(\text{in minutes}) = 8 * \text{LWR GE} + .06 * \text{TO}^2 - 5 * \text{LWRGE}^3 + 10.$$

Where LWRGE, decoding grade equivalent score, is used to determine the number of minutes of instruction a child should receive, and TO = fall LWR GE + 0.9 with a minimum GE of 2.0. This amount changes each month according to the following formula:

$$\text{CMMFRecc}(\text{in minutes}) = \text{CMMFAmount} - (8 - M / 0.7)$$

Where M, month, August = 0, September = 1, etc. The lowest amount recommended for CMMF is 5 minutes. This function is also non-linear and recommends small amounts of CMMF activities for students whose fall decoding score falls below 1.0, with rapidly increasing amounts for children reading at a grade equivalent falling between 1.0 (fall first grade) and 2.5. Very small amounts are recommended for students reading at a decoding grade equivalent above 2.75.

In previous investigations these formulas have been used to evaluate naturally occurring instruction, as well as provide recommendations to teachers enrolled in intervention studies. Once the prescribed amount of each type of instruction was determined for the fall and spring, the distance from this recommended amount of time was calculated using the coded observations. These distance from recommendation scores were then z-scored for further use in analyses.

The recommended amounts of time and observed amounts of time (in minutes) are listed in Table 1, averaged across all students and within the three categories of students observed in each classroom. As can be seen in this table, for all students the amount of teacher-managed code-focused instruction (TMCF) decreased over the school year while child-managed meaning-focused instruction (CMMF) increased. Both teacher-managed meaning-focused (TMMF) instruction and child-managed code-focused instruction (CMCF) decreased over the school year. In general, the most instructional time was spent on TMMF instruction. Correlations between the observed amount of time spent in each code and students' outcome scores are shown in Table 2.

Moving on from the creation of the distance from recommendation (DFR) scores, hierarchical linear modeling (HLM) software was used to model the relation between DFR scores at the fall and spring time points and student outcomes. Although all data used in these analyses were at the child-level, the amount of variation between classrooms indicated by the interclass correlation from the fully unconditional model (35.58% for decoding and 27.64% for passage comprehension) recommended the use of this type of modeling in order to account for the nested structure of our data with students nested within classrooms and schools. Identical models were created for both passage comprehension and decoding outcomes. In the initial model for both outcomes, the distance from recommendation of TMCF instruction in the fall was entered, and this additional variable reduced the variance explained in decoding by 0.43% and

reduced the variance explained in comprehension by 0.79%. The spring DFR for TMCF was then added, which explained an additional 0.32% of the variance in decoding and 0.21% of the variance in comprehension. Adding in DFR for CMMF in the fall accounted for 0.69% of the variance in decoding and 1.72% of the variance in comprehension. The final addition of DFR for CMMF in the spring accounted for 0.97% of the variance in decoding and 1.23% of the variance in comprehension. In total, the final model accounted for 2.41% of the variance in decoding (compared to the fully unconditional model) and 3.95% of the variance in comprehension (compared to the fully unconditional model). The final model evaluated was as follows:

$$Y_{ij} = [B_{0j} + r_{0ij}] + [B_{1j}(X_{1ij})] + [B_{2j}(X_{2ij})] + [B_{3j}(X_{3ij})] + [B_{4j}(X_{4ij})] + [B_{5j}(X_{5ij})] + e_{ij}$$

$$B_{0j} = \gamma_{00} + u_{0j}$$

$$B_{1j} = \gamma_{10}$$

$$B_{2j} = \gamma_{20}$$

$$B_{3j} = \gamma_{30}$$

$$B_{4j} = \gamma_{40}$$

$$B_{5j} = \gamma_{50}$$

Where Y_{ij} represents the spring literacy achievement score (decoding or comprehension) of child i in classroom j . X_{1ij} represents the fall literacy achievement score (decoding or comprehension) of child i in classroom j . X_{1ij} was centered around the grand mean. X_{2ij} and X_{3ij} represent the distance from recommendation of TMCF instruction in the fall and spring, respectively. X_{4ij} and X_{5ij} represent the distance from recommendation of CMMF instruction in the fall and spring, respectively.

Results from this model are shown in Tables 3 and 4; as can be seen in these tables, neither type of DFR score predicted either of the student outcomes.

Moving on from this model, a simpler model using only the observed amounts of TMCF and CMMF was used; again, identical models were created for both passage comprehension and decoding outcomes; in the fully unconditional decoding model, 35.58% of the variance was between classrooms, and in the fully unconditional comprehension model 27.64% of the variance was between models. The amount of TMCF instruction from the fall observation was added into the model, and this additional variable explained 1.15% of the variance in decoding and 0.98% of the variance in comprehension. The amount of TMCF observed in the spring was then added, which explained an additional 0.66% of the variance in decoding and 0.57% of the variance in comprehension. Adding CMMF from the fall observation accounted for 2.26% of the variance in decoding and 2.29% of the variance in comprehension. The final addition of CMMF in the spring accounted for 2.40% of the variance in decoding and 1.76% of the variance in comprehension. In total, the final model accounted for 6.37% of the variance in decoding (compared to the fully unconditional model) and 5.60% of the variance in comprehension (compared to the fully unconditional model).

$$Y_{ij} = [B_{0j} + r_{0ij}] + [B_{1j}(X_{1ij})] + [B_{2j}(X_{2ij})] + [B_{3j}(X_{3ij})] + [B_{4j}(X_{4ij})] + [B_{5j}(X_{5ij})] + e_{ij}$$

$$B_{0j} = \gamma_{00} + u_{0j}$$

$$B_{1j} = \gamma_{10}$$

$$B_{2j} = \gamma_{20}$$

$$B_{3j} = \gamma_{30}$$

$$B_{4j} = \gamma_{40}$$

$$B_{5j} = \gamma_{50}$$

Where Y_{ij} represents the spring literacy achievement score (decoding or comprehension) of child i in classroom j . X_{1ij} represents the fall literacy achievement score (decoding or comprehension)

of child i in classroom j . X_{1ij} was centered around the grand mean. X_{2ij} and X_{3ij} represent the observed amount of TMCF instruction in the fall and spring, respectively. X_{4ij} and X_{5ij} represent the observed amount of CMMF instruction in the fall and spring, respectively.

Results from this simpler model are shown in Tables 5 and 6; as can be seen in these tables, TMCF at the spring timepoint ($\beta = 0.016$, $p = 0.02$) and CMMF at the spring timepoint ($\beta = 0.012$, $p = 0.001$) both predicted student decoding outcome scores. This indicates that students who received more TMCF instruction during the spring observations had significantly higher decoding scores than their peers receiving less TMCF instruction during the same observation timepoint, with the same indication being made for those receiving more CMMF instruction during the spring observation timepoint. None of the four types of instruction predicted student passage comprehension outcomes scores.

Research Question 2: Relation of CLASS Scores to Reading Instruction and Student Achievement

CLASS domain scores in the fall and spring were shown to be stable over the school year, which is consistent with previous findings regarding the stability of the measure in the absence of specific coaching (Pianta, LaParo, & Hamre, 2008). For this reason, an average across the two observations of each domain was created and the descriptive information from this variable is shown in Table 7. Intercorrelations of the CLASS domains and dimensions are shown in Table 8; these correlations are similar to those from published studies using the tool in large, nationally representative datasets (Pianta, LaParo, & Hamre, 2008).

Pearson correlations were conducted to relate CLASS domain scores with Pathways/ISI instructional time codes, displayed in Table 9. These two measures were not significantly related at any point, indicating that they are capturing two distinct aspects of classroom instruction.

Hierarchical models were conducted by adding the CLASS domain scores in one at a time to the second (classroom) level of the following model described above:

$$Y_{ij} = [B_{0j} + r_{0ij}] + [B_{1j}(X_{1ij})] + [B_{2j}(X_{2ij})] + [B_{3j}(X_{3ij})] + [B_{4j}(X_{4ij})] + [B_{5j}(X_{5ij})] + e_{ij}$$

$$B_{0j} = \gamma_{00} + u_{0j}$$

$$B_{1j} = \gamma_{10} + [\gamma_{11}(N_{11j})] + [\gamma_{12}(N_{12j})] + [\gamma_{13}(N_{13j})] + u_{1j}$$

$$B_{2j} = \gamma_{20}$$

$$B_{3j} = \gamma_{30}$$

$$B_{4j} = \gamma_{40}$$

$$B_{5j} = \gamma_{50}$$

Where Y_{ij} represents the spring literacy achievement score (decoding or comprehension) of child i in classroom j . X_{1ij} represents the fall literacy achievement score (decoding or comprehension) of child i in classroom j . X_{1ij} was centered around the grand mean. N_{11j} , N_{12j} , and N_{13j} represent CLASS domain scores of emotional support, classroom organization, and instructional support, respectively. X_{2ij} and X_{3ij} represent the observed amount of TMCF instruction in the fall and spring, respectively. X_{4ij} and X_{5ij} represent the observed amount of CMMF instruction in the fall and spring, respectively. Entering in these variables accounted for an additional 4.32% of the variance in decoding and 2.17% of the variance in comprehension.

The results of this model are shown in Tables 10 and 11; with the addition of any of the CLASS domain scores, the previous significant relation between either TMCF or CMMF and decoding fell out. Classroom Organization ($\beta = 0.07$, $p = 0.060$) and Instructional Support ($\beta = 0.12$, $p = 0.023$) were predictive of decoding outcomes, whereas Emotional Support ($\beta = 0.14$, $p = 0.029$) and Classroom Organization ($\beta = 0.16$, $p = 0.045$) were predictive of student passage comprehension outcomes.

Research Question 3: The Role of Content Difficulty

Capturing the content used during classroom instruction proved a challenging task during the course of this study; although researchers had anticipated that teachers would use a number of worksheets, or teacher-created writing, in order to instruct students in literacy, these types of activities and materials were observed in less than a quarter of the classrooms observed. Indeed, only four teachers used worksheets at any time point, and all four of those teachers were within the same school. Similarly, only two teachers were observed presenting students with teacher-created writing (always in the form of completing a worksheet on the overhead projector); thus, analysis of this material was abandoned. The coding system created in anticipation of analyzing these worksheet materials was employed for the 12 worksheets collected. Worksheets provided on average 2.75 directions (Range=1 to 8), and had an average word frequency of 63.00 out of 88.3 (higher scores indicate simpler words). The worksheets contained on average 24.72 words total, and 15.71 unique words.

Because worksheets were not observed in more than 25% of the classrooms, and these classrooms were distinct and isolated from the rest of the sample, further analyses linking these materials with student outcomes were not conducted.

Collecting information regarding the texts used during classroom instruction was successful, with 290 books observed during the two observation timepoints. Of the 290 books students and teachers were observed engaging with, 238 were successfully given a Flesch-Kincaid score. Of the 52 books which were not scored, the majority (36) were student or class created books, and others were either out of print (10) or incorrectly noted in the observation narrative (6). Online databases and communication with publishers was used to collect Flesch-Kincaid scores for each book observed; in addition, the amount of time a child spent reading or

engaging with (independent reading or listening to teacher read aloud) each book was gathered from the written observation narratives. On average books that students were seen engaging with during the observation received a Flesch-Kincaid score of 2.99, and children spent on average 5.17 minutes reading or engaging with each book and a total of 17.89 minutes reading or engaging with books generally during the course of the observation.

An initial exploration of the relation between students' reading skills and the average difficulty level of the books they engaged in showed that fall reading skills and average book difficulty levels were unrelated ($\beta_{\text{FallLW}} = 0.035$, $p = 0.646$; $\beta_{\text{FallPC}} = 0.015$, $p = 0.824$); however, spring reading skills and average book difficulty levels were related ($\beta_{\text{SpringLW}} = 0.288$, $p = 0.003$; $\beta_{\text{SpringPC}} = 0.179$, $p = 0.008$). A difference score was created to show the difference between the students' reading skills and average book difficulty level in the fall and spring; graphs showing the distribution of these scores can be found in Figures 1-4. A lower score (on average, below 2.00) indicates an average book difficulty level falling 1.0 grade levels or more below the students' reading skills, whereas a high score (on average, above 3.00) indicates an average book difficulty level falling 1.0 grade levels or more above the students' reading skills. In general, the distributions of these difference scores are concentrated between 2.00 and 3.00 indicates an average difficulty level within 0.5 grade levels of the students' ability. These difference scores were then correlated with students' growth in reading achievement over the first grade year; these correlations are displayed in Table 12. The discrepancies between book difficulty and students' decoding and comprehension ability in the fall were moderately correlated with their growth in both decoding and comprehension over the school year. The discrepancy between book difficulty and students' decoding skills in the spring was related to growth in comprehension, but not decoding.

The Flesch-Kincaid score of the books observed was not significantly correlated with the amount of time spent in any type of instruction; further correlations between text-related information and time spent in the four types of instruction can be found in Table 13. Although these correlations were largely non-significant, the Pearson values were typically in the 0.20 to 0.40 range, and thus these data were used to further model children's achievement growth. The amount of time children spent with each book during the observation, and the amount of time students spent with books generally during the observation, were related to child passage comprehension and decoding outcomes in the following model described above:

$$Y_{ij} = [B_{0j} + r_{0ij}] + [B_{1j}(X_{1ij})] + [B_{2j}(X_{2ij})] + [B_{3j}(X_{3ij})] + [B_{4j}(X_{4ij})] + [B_{5j}(X_{5ij})] + e_{ij}$$

$$B_{0j} = \gamma_{00} + u_{0j}$$

$$B_{1j} = \gamma_{10}$$

$$B_{2j} = \gamma_{20}$$

$$B_{3j} = \gamma_{30}$$

$$B_{4j} = \gamma_{40}$$

$$B_{5j} = \gamma_{50}$$

Where Y_{ij} represents the spring literacy achievement score (decoding or comprehension) of child i in classroom j . X_{1ij} represents the fall literacy achievement score (decoding or comprehension) of child i in classroom j . X_{1ij} was centered around the grand mean. X_{2ij} and X_{3ij} represent the observed amount of time a child spent with each book in the fall and spring, respectively. X_{4ij} and X_{5ij} represent the observed amount of time a child spent with all books in the fall and spring, respectively.

The results of these models are displayed in Tables 14 and 15. The difficulty of the books students were seen interacting with during the observation was not significantly related to either

outcome. The amount of time children spent with each book was negatively related to both decoding ($\beta = -0.254, p < .01$) and passage comprehension ($\beta = -0.261, p < .01$) outcomes, indicating that students who spent more time on each book had lower scores on these measures at the spring timepoint than their peers who spent less time on each book. However, the total amount of time children were seen interacting with books was related to both decoding ($\beta = 0.152, p = 0.032$) and passage comprehension ($\beta = 0.11, p = 0.046$), indicating that students who spent more time interacting with books during the observation had higher scores on achievement tests at the spring time point than their peers who spent less time interacting with books.

Research Question 4: Linking Teacher Ratings of Student Ability with Assessments and Instruction

Teacher ratings from the fall and spring were compared with student scores on standardized achievement tests at the same timepoints. First, the ratings with correlated with test scores as shown in Table 16. Teacher ratings were significantly related to both student achievement in both decoding and comprehension. These ratings were also correlated with both Pathways/ISI codes TMCF and CMMF, as well as all three CLASS domains, also displayed in Table 16. Because these ratings were not significantly correlated with observations of classroom instruction, further analyses using these data were not conducted.

Research Question 5: The Role of Teacher Experience

The final set of models evaluated for this study included measures of teachers' years of teaching experience. The teaching experience of the teachers enrolled in the study varied greatly, both across all grades taught ($M=15.54$ years, Range=4-38 years) and specifically within first grade classrooms ($M=8.39$ years, Range=1.5-20 years). This aspect of teachers' characteristics

proved to be more diverse than teachers' education, which was relatively consistent across all teachers.

The years of experience teachers had spent teaching first grade specifically was evaluated as using the following model for both decoding and passage comprehension outcomes:

$$Y_{ij} = [B_{0j} + r_{0ij}] + [B_{1j}(X_{1ij})] + [B_{2j}(X_{2ij})] + [B_{3j}(X_{3ij})] + [B_{4j}(X_{4ij})] + [B_{5j}(X_{5ij})] + e_{ij}$$

$$B_{0j} = \gamma_{00} + u_{0j}$$

$$B_{1j} = \gamma_{10}$$

$$B_{2j} = \gamma_{20} + [\gamma_{21}(N_{21j})] + u_{2j}$$

$$B_{3j} = \gamma_{30} + [\gamma_{31}(N_{31j})] + u_{3j}$$

$$B_{4j} = \gamma_{40} + [\gamma_{41}(N_{41j})] + u_{4j}$$

$$B_{5j} = \gamma_{50} + [\gamma_{51}(N_{51j})] + u_{5j}$$

Where Y_{ij} represents the spring literacy achievement score (decoding or comprehension) of child i in classroom j . X_{1ij} represents the fall literacy achievement score (decoding or comprehension) of child i in classroom j . X_{1ij} was centered around the grand mean. N_{X1j} represents teachers' years of experience teaching first grade. X_{2ij} and X_{3ij} represent the observed amount of TMCF instruction in the fall and spring, respectively. X_{4ij} and X_{5ij} represent the observed amount of CMMF instruction in the fall and spring, respectively.

Model results are contained in Tables 17 and 18; years of teaching first grade was predictive of the relation between TMCF and both decoding ($\beta=0.002$, $p=0.09$) and passage comprehension ($\beta=0.006$, $p=0.04$) outcomes. This relation indicates that students who receive more teacher-managed code-focused instruction have higher reading achievement scores, and that this relation is more pronounced for students of teachers with more years of experience teaching first grade.

Chapter 5. Discussion

In this section, the results for each research question will be interpreted, and the study as a whole will be reviewed. Finally, next steps and conclusions which can be drawn from this study will be considered.

Research Question 1: Individualization of Instruction

In general, students enrolled in the study represented a diverse array of first graders in terms of their reading abilities. Students from each of our categories (above-, below-, and at-grade level) were easily identifiable based on their initial reading assessment scores. Thus, we could reasonably expect that the established predictions from the Pathways/ISI system would be appropriate for this group of students. Naturally occurring instruction did not always meet these recommendations, which was to be expected in a study of teachers who had not been exposed to this prior research and findings. Indeed, the use of these recommendations was done post hoc, and teachers were not informed about the coding systems that would be used or the specifics of the analyses that would be done with coded instruction data. The instruction observed did, however, follow patterns that the Pathways/ISI system would recommend generally; that is, the amount of TMCF decreased over the school year while the amount of CMMF increased. From the initial studies conducted by Morrison, Connor and colleagues, this type of change over time was linked to student reading ability growth (Connor, Morrison, & Slominski, 2006).

Analyses using the DFR scores created with the observed and recommended instruction from across the school year were unrelated to student's reading ability at outcome. It may be that the formula used to predict the amount and type of instruction students needed was not appropriately calibrated for this group of students; the formula limits growth to one grade level, which was originally intended to assist teachers trying to get struggling readers up to grade level.

Indeed, the initial studies using the formalized Pathways/ISI system were comprised largely of students who began the year reading below grade level (Connor et. Al, 2008). Although some students enrolled in this study did enter first grade with reading abilities measured at kindergarten and lower, they did not constitute the majority of the student population. Potentially creating a formula which left growth unhindered or restricted growth to more than one grade level might be an appropriate direction for future analyses with this data.

This finding was unexpected and disappointing, but suggested that perhaps the amount and type of instruction teachers provided naturally might be the best predictor of students' outcome scores. Using the amount of TMCF and CMMF instruction in models with decoding and passage comprehension student outcomes demonstrated that the naturally occurring instruction observed in classrooms was significantly related to student performance on reading tests at the end of the school year. Both types of instruction were predictive of student performance in decoding and passage comprehension, indicating that teachers were providing students with useful skills in both code- and meaning-related concepts during classroom instructional time.

The pattern of changes in the amount of instructional time across the school year, as well as the links between naturally occurring instruction and student reading achievement, both indicate that teachers were individualizing instruction in an effective way. Further research with this dataset could improve the Pathways/ISI recommendation formulas. Moreover, these findings indicate that the live coding system was able to successfully capture important aspects of the classroom environment. An important next step in determining the applications of this live coding scheme would be to recode activities from the running narratives, and explore ways in which more specific activities, or groupings, might also be related to student outcomes.

Research Question 2: Relation of CLASS Scores to Reading Instruction and Student Achievement

Findings related to the question of how CLASS scores are linked to reading instruction and student achievement followed some of the patterns observed in previous studies comparing Pathways/ISI with CLASS. Specifically, the two measures were not significantly correlated with one another. This particular finding has been proposed to indicate that the two measures capture different aspects of the classroom experience for students (Guthrie, Grammer, & Morrison, 2013). Prior studies have suggested that the two measures are complementary; this supposition was not borne out by the findings from this study. When CLASS scores were entered into models of student achievement as predicted by Pathways/ISI codes, the effects of Pathways/ISI amount and type of instruction dropped out of the model; indeed, the amount of variance explained by models with Pathways/ISI codes when CLASS codes were included in the dataset were negative.

However, CLASS scores were related to the degree of growth students experienced from fall to spring of the first grade year. Specifically, the organization and orderliness of the classroom as well as the amount of instructional support provided in the classroom were related to students' decoding skill growth. In addition, the warmth and responsivity present in student-teacher interactions and the degree of organization of the classroom environment were related to students' comprehension skill growth.

Studies which use exclusively the CLASS measure to predict student achievement have found relations between all three domains of the CLASS with reading and mathematics achievement. Typically, classrooms which have more instructional support (e.g., advanced language, higher order questions) are the most effective at improving student achievement.

However, the most consistent finding is that emotional support (e.g., warmth, engagement) is strongly linked with student achievement growth. It is not entirely clear why some domains of the CLASS would be related to our student outcome scores of interest. Organization, which is linked to both outcomes, measures teacher preparedness as well as student productivity. It seems clear that both would be beneficial to students across a broad array of learning outcomes. It is more difficult to determine why instructional support or emotional support would not be consistently related to both outcomes given the evidence from prior studies.

The findings from this particular research question seem to indicate that although CLASS scores are consistent across time and relatively robust to small sample size, the live coded Pathways/ISI system is not as robust. CLASS coding is intended to be done at the level of the average child experience, and is meant to be a live-coded system. But the limitations of the Pathways/ISI system (namely, limiting the number of children observed) and the method in which it was applied to classroom observations (children of three specific ability levels were observed) did not produce a consistent evaluation of instructional effectiveness. Continuing to alter and improve the live coding system is an important next step; in addition, using the traditional Pathways/ISI coding scheme with the CLASS should be done in larger and diverse datasets with the goal of determining the interrelations of these two measures.

Research Question 3: The Role of Content Difficulty

To explore this question, analyses delved further into the study of how teachers may be individualizing instruction in ways not previously captured by the Pathways/ISI system. The original focus of this study stemmed from a need to explore why some types of instruction were not effective for all children, and taking that idea a step further, also exploring how those types of instruction might be made more effective for all students. It was posited that materials and the

content of lessons might be a method teachers would be using to differentiate instruction between students, and that this method would not be observed through the original Pathways/ISI system. Unfortunately, the teachers enrolled in this study did not utilize many of the materials anticipated, which was an interesting finding in and of itself. The materials collected were all worksheets from student workbooks; indeed, both the teacher-created writing observed and the physical materials distributed consisted of phonics-based worksheet activities. Because these kinds of materials were only observed in at most four classrooms, it did not seem appropriate to use the information gleaned from these materials to predict student outcomes.

The texts that teachers and students selected for reading, and specifically the time that students were engaged with those texts, proved a significant finding of this study. Teachers and students were observed using a broad array of texts, from predictable text books (e.g., *Brown Bear, Brown Bear*) to nonfiction works (e.g., *Encyclopedia of Dinosaurs*) to chapter books (e.g., *Diary of a Wimpy Kid*). This variety led to a wide range of Flesch-Kincaid reading ease scores, which were unfortunately not predictive of student outcomes in the models analyzed. It is possible that this finding is related to the differences in teacher approaches to student book selection. Some teachers allowed students free choice in selecting a book, sometimes after training students in how to pick a “just right” book which would challenge them. Other teachers had prescribed (based on reading performance during teacher assessment) bins or sections of books from which students could borrow. Still other teachers did not provide books for students to choose between, and instead allowed students to interact with curriculum created books during direct instructional time only. This led to large differences in the books children were able to interact with, and this difference between classrooms may have accounted for most of the differences in student reading ability as related to book difficult. For example, teachers who

allowed free choice may have had a classroom of students with higher reading motivation or skills than the teachers who did not provide supplemental texts.

Some specific analyses comparing the average book difficulty level with students' reading skills showed a linear relation between reading skills and book difficulty in the spring, but not in the fall. This could support anecdotal evidence that teachers would have limited students' book use to "just right" texts as the year went on and they gained more information about students' reading skills. In addition, students were typically observed reading books within half a grade level of their current reading ability at each observation. Interestingly, correlations between students' growth in reading skills and the difference between students reading skills and average book difficulty level were significant for all fall difference scores, but only for one of the spring difference scores. This could provide some information about whether or not teachers limiting students' book choices is a good strategy for high achieving students. These correlations indicate that students who read more challenging texts grow more in reading skills than those students who read less challenging texts. More research within this data about the effects of "just right" books, or those books near or just above students' reading skills, versus challenging or easy books is needed in order to continue identifying content as an aspect of effective instruction.

Two additional variables regarding students' engagement with texts were created using the information gathered from the coding narratives. Specifically, the average amount of time a student spent with a single book and the amount of time a student spent with all books were both used to predict student outcomes on decoding and comprehension. The average amount of time a student spent with a single book was negatively related to students' end of year scores in decoding and comprehension. This variable may have also captured which students were

struggling versus advanced readers; for example, a struggling reader may spend more time with each book due to difficulty decoding and/or comprehending the text.

The total amount of time students spent engaged with books was positively related to both decoding and comprehension outcome scores. This finding relates well to previous research regarding ‘eyes on text’ (e.g., Brenner & Hiebert, 2010) which indicates that students benefit from reading instruction which allows them to have their eyes on the text or book that is the focus of the current lesson. This particular variable was considered distinct from CMMF instruction time, which was also significantly predictive of these outcomes, because of the amount of independent writing/journaling time that students were observed to engage in, as well as the fact that this measure includes both teacher- and child-managed meaning-focused activities. Thus, this finding provides unique insight into a component of effective instruction.

Further considerations using alternative measures of text readability would be worthwhile next steps with this data; specifically, continuing to explore more nuanced measures of the readability of early texts (as described by Hiebert & Pearson, 2010) might result in links between text difficulty and child outcomes. In addition, potentially using these measures, along with instructional coding schemes, in larger datasets with high quality video observations of classrooms could help create a more robust model of the connection between instruction, texts, and students’ reading skill growth.

Research Question 4: Linking Teacher Ratings of Student Ability with Assessments and Instruction

Questionnaires regarding numerous reading-related abilities were given to each teacher for all of their students enrolled in the study. Teachers proved to be adept at rating the abilities of their students using the scale provided, as their ratings were significantly related to student

ability. It is interesting to note that, despite the ratings of student abilities being relatively accurate, these ratings were not significantly related to either the Pathways/ISI instructional codes or classroom scores on CLASS domains, despite the fact that both of these measures were related to student achievement scores. Because the questionnaire provided five broad categories of ability for teachers to place students in, it seems likely that the nuances detectable in the broader sample of students was not present in the questionnaire data. In addition, the sample of students available for comparisons between questionnaire and observation was much smaller, and small relations were likely undetectable. Because of the accuracy of the ratings, it would be useful to continue to develop such questionnaires to examine the ways in which teachers' perceptions of students' skills naturally link with their instruction, and how these links can be improved for individualized instruction.

Research Question 5: The Role of Teacher Experience

Previous studies provided strong evidence for the role of teacher experience, above and beyond teacher education, certification, and other characteristics, in highly effective instruction (e.g., Huang & Moon, 2009; Guthrie, Connor, & Morrison, 2010). The data from this study bore out this relation; indeed, although the years a teacher had taught at any grade level was not predicted to students' growth in decoding or comprehension, teachers' experience teaching first grade were significantly related to students growth on both measures of reading ability. Interpretation of this model reveals that students benefited in terms of both decoding and comprehension ability from more time spent in TMCF and CMMF instruction, and that students of more experience first grade teachers received an extra boost in reading ability on both measures.

Previous work looking at more experienced or even master teachers has revealed that they may be more adept at presenting instructional material. More research within this dataset and others should explore the processes underlying the steeper slope in growth students of more experienced teachers. Are these teachers providing more precisely individualized instruction? Are they spending less time in non-instructional activities, such as transitions?

Limitations

One of the most difficult limitations for this study to overcome was the number of students observed at each timepoint. Although many students within each classroom were enrolled in the study and assessed ($M=15$, $\text{Range}=10-23$), only three students from each classroom were observed during instruction ($N=48$). This limited the findings which could be gleaned from models in which the classroom observations were entered, and may have allowed for the detection of only very large effects. The adapted coding system did work well, and the ease of using this system certainly recommends it for future studies; however, the ability to code more students is a vital contribution from the original Pathways/ISI coding method.

For a study of the unique role of content in first grade literacy instruction, more materials needed to be collected in order to adequately describe this aspect of the classroom. The reason for this lack of materials has two obvious possible explanations; first, teachers may have avoided giving students worksheets during observations, preferring to have the observation cover a particular group of activities. Second, a large group of the teachers in two of the schools used a unique set of daily literacy activities for their morning literacy block, which was typically the time of observation. This set of activities included students' independent reading at three separate stations, students' independent journaling/story-writing at one station, and students' independent phonemic awareness and letter-sound activities at the last station. Students were

pulled out for teacher-led small groups, which appeared to be ability-based. The activities included in this method of instruction did not call for or truly allow for worksheets; even the word-work activity station involved finding letters, sounds, and rhymes in books or around the classroom.

In addition to this content-based limitation, there were also barriers to collecting all of the possible information regarding the Flesch-Kincaid reading-ease scores of the texts observed. Of the 290 books students and teachers were observed engaging with, 238 were successfully given a Flesch-Kincaid score. Of the 52 books that were not scored, the majority (36) were student or class created books, and others were either out of print (10) or incorrectly noted in the observation narrative (6).

Implications

The first implication from the relatively limited findings of this study indicates that more work needs to be done on the feasibility of creating a universal set of recommendations for teachers regarding the appropriate amounts and type of instructional activities to provide to students based on reading ability. The lack of child-by-instruction interaction findings in this data might be related to the generally high scores students from this sample received on reading skill assessments. Although students were drawn from diverse public schools, the average decoding and comprehension scores of the sample at baseline and outcome were above grade-level. Thus, it seems possible that the needs of these high-ability students were not reflected in the recommendation equations created from lower-ability samples.

The connection between the CLASS, Pathways/ISI, and other classroom coding systems is still inconsistent. Although the two measures have been consistently unrelated when compared using Pearson correlations, this study found that entering CLASS variables into models

predicting literacy achievement growth resulted in nonsignificant findings regarding the relation between instructional time and student outcomes, which has not been the case in prior work (e.g., Guthrie, Grammer, & Morrison, 2013). A study in which these two systems, along with other content-general and content-specific observation systems, are used to explore large samples of students enrolled in early elementary classrooms would not only help to tease apart these components of the classroom, but also help in creating a more comprehensive measure which captures multiple dimensions of instruction which are related to effective teaching.

The findings regarding the amount of time students spent reading books is promising, and reflects similar findings from the Eyes on Text literature. Using alternative measures of text readability, and attempting to gather information about texts used in video observations from larger datasets, might help to explore why text difficulty was not related to student outcomes, and how time spent with books fits into the Pathways/ISI meaning-focused instructional codes.

Future research should continue to examine the nuances of how teachers go about individualizing instruction for students, and how their reading instruction is related to larger classroom variables. Teachers were able in this study to accurately rate their students' abilities within the first few weeks of school, although these ratings were not significantly related to observed instruction or classroom climate. However, student abilities were predicted by the amount and time of instruction that students received. Thus, ratings which are on a larger, more specific scale may allow researchers to explore the relation between teacher perceptions and actions in greater detail. Determining the pathway between teachers' knowledge of student skill level and their instructional decisions would be beneficial not only for identifying effective teachers, but also for nominating teachers who might benefit from additional professional development or specific tools.

Conclusion

This study provides an interesting initial look into not only the role of content in early literacy instruction, but also the prevalence of materials and texts used in these classrooms. The amount of time students spent in specific types of instruction did predict student outcomes, and broader conceptualizations of the classroom climate were also related to student reading ability growth. Findings from this study provided more evidence for the importance of students spending time interacting with text, and also support previous research regarding the importance of teachers' years of experience teaching their current grade level. Overall, this study provided a first step towards creating a comprehensive conceptualization of the aspects of the classroom, teacher, and students related to effective instruction.

Table 1. Average observed instruction and recommended instruction from the fall and spring timepoints in each of the four Pathways/ISI codes, separated by student reading ability.

Pathways/ISI by Ability												
	TMCF				CMMF				TMMF		CMCF	
	<u>Fall</u> <u>Obs</u>	<u>Fall</u> <u>Recc</u>	<u>Spring</u> <u>Obs</u>	<u>Spring</u> <u>Recc</u>	<u>Fall</u> <u>Obs</u>	<u>Fall</u> <u>Recc</u>	<u>Spring</u> <u>Obs</u>	<u>Spring</u> <u>Recc</u>	<u>Fall</u> <u>Obs</u>	<u>Spring</u> <u>Obs</u>	<u>Fall</u> <u>Obs</u>	<u>Spring</u> <u>Obs</u>
Below GL	15.53	6.53	11.74	5.58	13.02	13.41	19.37	20.55	27.57	25.76	7.77	3.4
At GL	16.48	5.27	11.25	4.36	21.31	14.72	18.1	21.26	27.25	25.04	7.21	4.31
Above GL	13.17	4.04	11.67	3.89	19.61	15.93	21.37	23.07	28.86	23.17	5.1	6.44
Overall	15.05	4.33	11.55	4.12	18.09	15.14	19.62	22.29	27.9	24.63	6.67	4.75

Table 2. Correlations between Pathways/ISI codes and students' achievement in decoding, comprehension, and vocabulary in the fall and spring.

	Correlations					
	Fall Decoding	Fall Comprehension	Fall Vocabulary	Spring Decoding	Spring Comprehension	Spring Vocabulary
<u>Fall TMMF</u>	0.003	-0.041	0.138*	-0.007	0.002	-0.071
<u>Fall TMCF</u>	-0.001	-0.025	0.121	0.046	-0.013	0.02
<u>Fall CMMF</u>	0.211**	0.189**	0.124	0.07	0.121	0.118
<u>Fall CMCF</u>	0.085	0.088	0.012	0.165	-0.105	0.073
<u>Spring TMMF</u>	-0.021	-0.062	0.101	0.033	0.068	0.103
<u>Spring TMCF</u>	0.137	-0.086	0.119	0.132	0.012	-0.087
<u>Spring CMMF</u>	0.044	0.142*	0.191**	0.036	0.128	0.170*
<u>Spring CMCF</u>	0.211	0.013	0.131	0.140*	-0.002	0.014

Table 3. Model results using distance from recommendation (DFR) scores in each of the Pathways/ISI codes to predict students' growth in decoding over the school year.

Estimated Effects					
	Beta Coefficients	Standard Error	t-Statistics	p-Value	df
Intercept	15.31	3.85	3.981	0.002	32
Fall Decoding	2.984	0.984	2.801	0.002	28
Fall DFR TMCf	0.531	0.382	1.392	0.166	28
Spring DFR TMCf	0.013	0.076	0.168	0.867	28
Fall DFR CMMF	-0.260	0.118	-0.352	0.213	28
Spring DFR CMMF	-0.010	0.036	-0.289	0.773	28
Chi Squares					
Parameter	Standard Deviation	Variance	Degrees of Freedom	Chi- Square	p- Value
Intercept	14.840	220.223	26	2206.705	0.001

Table 4. Model results using distance from recommendation (DFR) scores in each of the Pathways/ISI codes to predict students' growth in comprehension over the school year.

Estimated Effects						
	Beta	Standard	t-Statistics	p-Value	df	
	Coefficients	Error				
Intercept	27.315	2.620	3.981	0.002	32	
Fall Comprehension	3.995	0.930	2.204	0.001	28	
Fall DFR TMCF	-0.297	0.410	-0.731	0.466	28	
Spring DFR TMCF	-0.093	0.078	-1.184	0.238	28	
Fall DFR CMMF	0.159	0.149	1.067	0.288	28	
Spring DFR CMMF	0.036	0.034	1.067	0.288	28	
Chi Squares						
Parameter	Standard	Variance	Degrees of	Chi-	p-	
	Deviation		Freedom	Square	Value	
Intercept	14.840	220.223	26	2206.705	0.001	

Table 5. Model results using the amount of observed time spent in each of the Pathways/ISI codes to predict students' growth in decoding over the school year.

Estimated Effects					
	Beta Coefficients	Standard Error	t-Statistics	p-Value	df
Intercept	110.917	1.950	56.877	0.001	32
Fall Decoding	0.558	0.117	4.764	0.001	28
Fall Observed TCMF	-0.002	0.004	-0.499	0.621	28
Spring Observed TCMF	0.016	0.006	2.562	0.016	28
Fall Observed CMMF	0.003	0.004	0.787	0.438	28
Spring Observed CMMF	0.013	0.003	3.606	0.001	28
Chi Squares					
Parameter	Standard Deviation	Variance	Degrees of Freedom	Chi- Square	p- Value
Intercept	10.817	117.015	26	409.050	0.001

Table 6. Model results using the amount of observed time spent in each of the Pathways/ISI codes to predict students' growth in comprehension over the school year.

Estimated Effects						
	Beta	Standard				
	Coefficients	Error	t-Statistics	p-Value	df	
Intercept	103.465	1.867	55.429	0.001	32	
Fall Comprehension	1.172	0.281	4.159	0.001	28	
Fall Observed TMCF	-0.004	0.007	-0.586	0.562	28	
Spring Observed TMCF	0.017	0.003	2.136	0.025	28	
Fall Observed CMMF	0.002	0.005	0.720	0.477	28	
Spring Observed CMMF	0.012	0.006	2.336	0.027	28	

Chi Squares					
Parameter	Standard		Degrees of	Chi-	p-
	Deviation	Variance	Freedom	Square	Value
Intercept	9.98	99.762	26	233.654	0.001

Table 7. Average CLASS Domain scores

Average CLASS Domain Scores			
	<u>Emo Supp</u>	<u>Class Org</u>	<u>Inst Supp</u>
Mean	5.86	5.36	2.42
SD	0.8	0.63	0.92
Range	3.25-6.75	3.84-6.67	1.00-5.17

Table 8. Correlations amongst CLASS domain and dimension scores.

	Correlations												
	PC	NC	TS	RSP	BM	PD	ILF	CD	QLF	LM	Emo Supp	Class Org	Inst Supp
<u>Positive Climate</u>	--	.045	.869**	.840**	.794**	.762**	.741**	.610*	.551*	.752**	.927**	.852**	.691**
<u>Negative Climate</u>	.045	--	.175	.015	.099	-.034	-.281	.051	-.014	.293	.169	-.104	.104
<u>Teacher Sensitivity</u>	.869**	.175	--	.857**	.764**	.664**	.663**	.486	.453	.732**	.949**	.772**	.594*
<u>Respect for Student Perspective</u>	.840**	.015	.857**	--	.642**	.671**	.668**	.671**	.579*	.784**	.870**	.739**	.736**
<u>Behavior Management Productivity</u>	.794**	.099	.764**	.642**	--	.788**	.696**	.609*	.549*	.632*	.793**	.906**	.655**
<u>Instructional Learning Formats</u>	.762**	-.034	.664**	.671**	.788**	--	.640*	.595*	.442	.700**	.715**	.893**	.627*
<u>Instructional Learning Formats</u>	.741**	-.281	.663**	.668**	.696**	.640*	--	.557*	.600*	.685**	.647**	.890**	.666**
<u>Cognitive Development Quality of Feedback</u>	.610*	.051	.486	.671**	.609*	.595*	.557*	--	.779**	.765**	.563*	.652**	.946**
<u>Quality of Feedback</u>	.551*	-.014	.453	.579*	.549*	.442	.600*	.779**	--	.658**	.429	.597*	.903**
<u>Language Modeling</u>	.752**	.293	.732**	.784**	.632*	.700**	.685**	.765**	.658**	--	.751**	.754**	.869**

Table 9. Correlations between CLASS scores and instructional variables in the fall and spring.

	Fall TCMF	Fall CMMF	Spring TCMF	Spring CMMF
<u>Emo Supp</u>	0.19	-0.08	0.21	0.18
<u>Class Org</u>	0.14	0.08	0.17	-0.07
<u>Inst Supp</u>	-0.01	0.01	0.08	0.15

Table 10. Model results using the amount of observed time spent in each of the Pathways/ISI codes to predict students' growth in decoding over the school year, with CLASS Domain scores predicting students' growth on decoding.

Estimated Effects					
	Beta	Standard			
	Coefficients	Error	t-Statistics	p-Value	df
Intercept	106.110	0.873	121.480	0.001	10
Fall Decoding	0.747	0.144	5.201	0.001	8
Emotional Support	0.144	0.052	2.744	0.029	
Classroom Organization	0.161	0.066	2.427	0.045	
Fall Observed TCMF	-0.004	0.007	-0.586	0.562	8
Spring Observed TCMF	0.017	0.003	2.136	0.025	8
Fall Observed CMMF	0.002	0.005	0.720	0.477	8
Spring Observed CMMF	0.012	0.006	2.336	0.027	8
Chi Squares					
Parameter	Standard		Degrees of	Chi-	p-
	Deviation	Variance	Freedom	Square	Value
Intercept	2.048	4.193	10	25.563	0.005

Table 11. Model results using the amount of observed time spent in each of the Pathways/ISI codes to predict students' growth in comprehension over the school year, with CLASS Domain scores predicting students' growth on comprehension.

Estimated Effects					
	Beta	Standard			
	Coefficients	Error	t-Statistics	p-Value	df
Intercept	114.130	0.760	150.134	0.001	10
Fall Comprehension	0.754	0.077	9.735	0.001	8
Classroom Organization	0.071	0.033	2.116	0.067	
Instructional Support	0.118	0.042	2.816	0.023	
Fall Observed TCMF	-0.004	0.007	-0.586	0.562	8
Spring Observed TCMF	0.017	0.003	2.136	0.025	8
Fall Observed CMMF	0.002	0.005	0.720	0.477	8
Spring Observed CMMF	0.012	0.006	2.336	0.027	8
Chi Squares					
Parameter	Standard		Degrees of	Chi-	p-
	Deviation	Variance	Freedom	Square	Value
Intercept	2.249	5.059	14	22.130	0.014

Table 12. Correlations between student achievement and book readability discrepancy scores and students' growth over the school year in reading skills.

	Fall Decoding Discrepancy	Fall Comprehension Discrepancy	Spring Decoding Discrepancy	Spring Comprehension Discrepancy
<u>Decoding Growth</u>	0.36**	0.30**	-0.04	0.01
<u>Comprehension Growth</u>	0.38**	.49**	0.26**	0.12

Table 13. Correlations between text-related data (book difficulty, total number of books read, and time spent reading books) and all four types of literacy instruction in the spring and fall.

Correlations						
	Fall Book Difficulty	Fall Number of Books	Fall Time with Books	Spring Book Difficulty	Spring Number of Books	Spring Time with Books
<u>Fall TMMF</u>	0.091	0.309	0.253	0.077	0.11	0.273
<u>Fall TMCF</u>	0.007	-0.221	-0.31	0.076	0.056	-0.226
<u>Fall CMMF</u>	0.096	0.255	0.243	-0.036	0.237	0.345
<u>Fall CMCF</u>	0.034	-0.226	-0.287	0.101	0.058	-0.251
<u>Spring TMMF</u>	0.093	0.263	0.198	0.036	0.215	0.341
<u>Spring TMCF</u>	0.076	-0.366	-0.219	0.056	-0.13	-0.309
<u>Spring CMMF</u>	0.027	0.251	0.207	0.058	0.408	.613*
<u>Spring CMCF</u>	0.039	-0.376	-0.295	-0.002	-0.116	-0.263

Table 14. Model results using the amount of time students spent reading books during the observation, as well the amount of time they spent engaging with each individual book, to predict students' growth in decoding over the school year.

Estimated Effects					
	Beta	Standard			
	Coefficients	Error	t-Statistics	p-Value	df
Intercept	112.913	1.981	55.998	0.001	32
Fall Decoding	0.558	0.258	2.170	0.039	28
Time Per Book	-0.254	0.010	1.614	.001	28
Total Time Reading	0.152	0.118	0.352	0.32	28
Chi Squares					
Parameter	Standard		Degrees of	Chi-	p-
	Deviation	Variance	Freedom	Square	Value
Intercept	3.499	1.225	26	7.267	0.500

Table 15. Model results using the amount of time students spent reading books during the observation, as well the amount of time they spent engaging with each individual book, to predict students' growth in comprehension over the school year.

Estimated Effects						
	Beta	Standard				
	Coefficients	Error	t-Statistics	p-Value	df	
Intercept	107.096	0.500	214.040	0.001	10	
Fall Comprehension	0.453	0.083	5.462	0.001	8	
Time Per Book	-0.261	0.052	4.012	0.001	8	
Total Time Reading	0.110	0.006	3.016	0.046	8	
Chi Squares						
Parameter	Standard		Degrees of	Chi-	p-	
	Deviation	Variance	Freedom	Square	Value	
Intercept	3.499	1.225	10	7.267	0.500	

Table 16. Teacher ratings of students' literacy skills correlated with students' concurrent performance on reading assessments, as well as both CLASS and Pathways/ISI observation codes.

Correlations							
	Decoding	Comprehension	TMCF	CMMF	Emotional Support	Classroom Organization	Instructional Support
<u>Letter Knowledge Rating</u>	0.69**	0.69**	-0.07	0.03	0.11	0.03	0.04
<u>Phonemic Awareness Rating</u>	0.75**	0.75**	0.04	-0.18	-0.13	0.07	0.11
<u>Comprehension Rating</u>	0.73**	0.73**	-0.19	0.13	-0.01	0.19	0.06

Table 17. Model results using the amount of observed time spent in each of the Pathways/ISI codes to predict students' growth in decoding over the school year, with teacher's years of experience teaching first grade predicting the degree of growth in decoding.

Estimated Effects					
	Beta	Standard			
	Coefficients	Error	t-Statistics	p-Value	df
Intercept	114.918	0.498	230.591	0.001	10
Fall Decoding	0.897	0.038	23.811	0.001	8
Grade-Level Experience	0.002	0.001	2.023	0.097	
Fall Observed TCMF	0.002	0.001	-1.380	0.226	8
Spring Observed TCMF	0.002	0.001	2.289	0.069	8
Fall Observed CMMF	0.001	0.001	0.620	0.562	8
Spring Observed CMMF	0.004	0.001	3.796	0.018	8
Chi Squares					
Parameter	Standard		Degrees of	Chi-	p-
	Deviation	Variance	Freedom	Square	Value
Intercept	2.626	0.690	10	9.277	0.235

Table 18. Model results using the amount of observed time spent in each of the Pathways/ISI codes to predict students' growth in comprehension over the school year, with teacher's years of experience teaching first grade predicting the degree of growth in comprehension.

Estimated Effects					
	Beta	Standard			
	Coefficients	Error	t-Statistics	p-Value	df
Intercept	107.13	0.513	208.505	0.001	10
Fall Comprehension	0.544	0.118	4.615	0.005	8
Grade-Level Experience	0.007	0.002	2.755	0.040	
Fall Observed TCMF	-0.007	0.004	-1.708	0.148	8
Spring Observed TCMF	0.003	0.002	1.201	0.284	8
Fall Observed CMMF	0.004	0.005	0.775	0.474	8
Spring Observed CMMF	0.026	0.002	1.326	0.242	8
Chi Squares					
Parameter	Standard		Degrees of	Chi-	p-
	Deviation	Variance	Freedom	Square	Value
Intercept	2.583	0.667	10	7.270	0.500

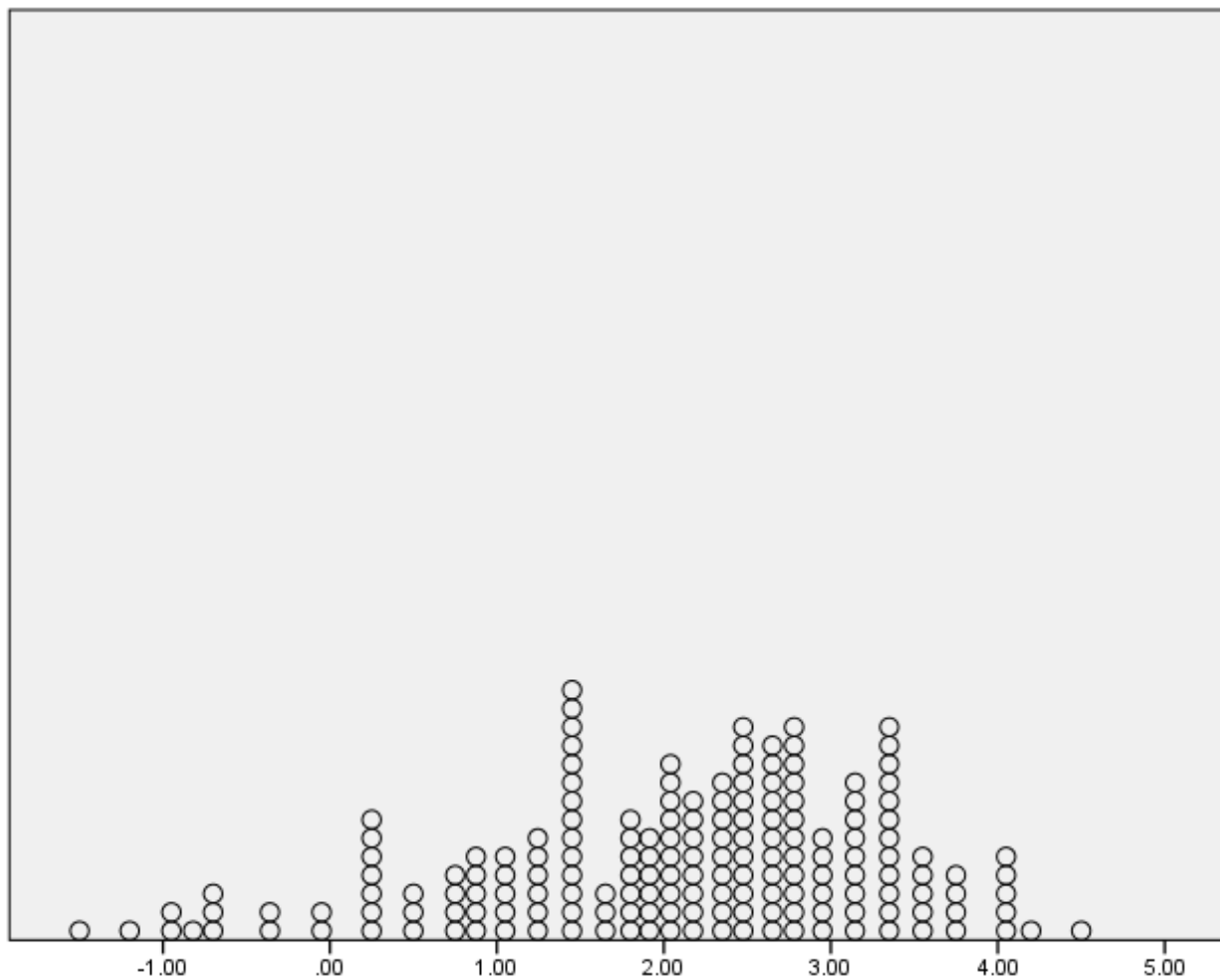
Figures

Figure 1. Difference between students' scores on decoding in the fall and the average difficulty level of books interacted with in the fall; each dot represents one student, and the scale on the X-axis is in grade-levels.

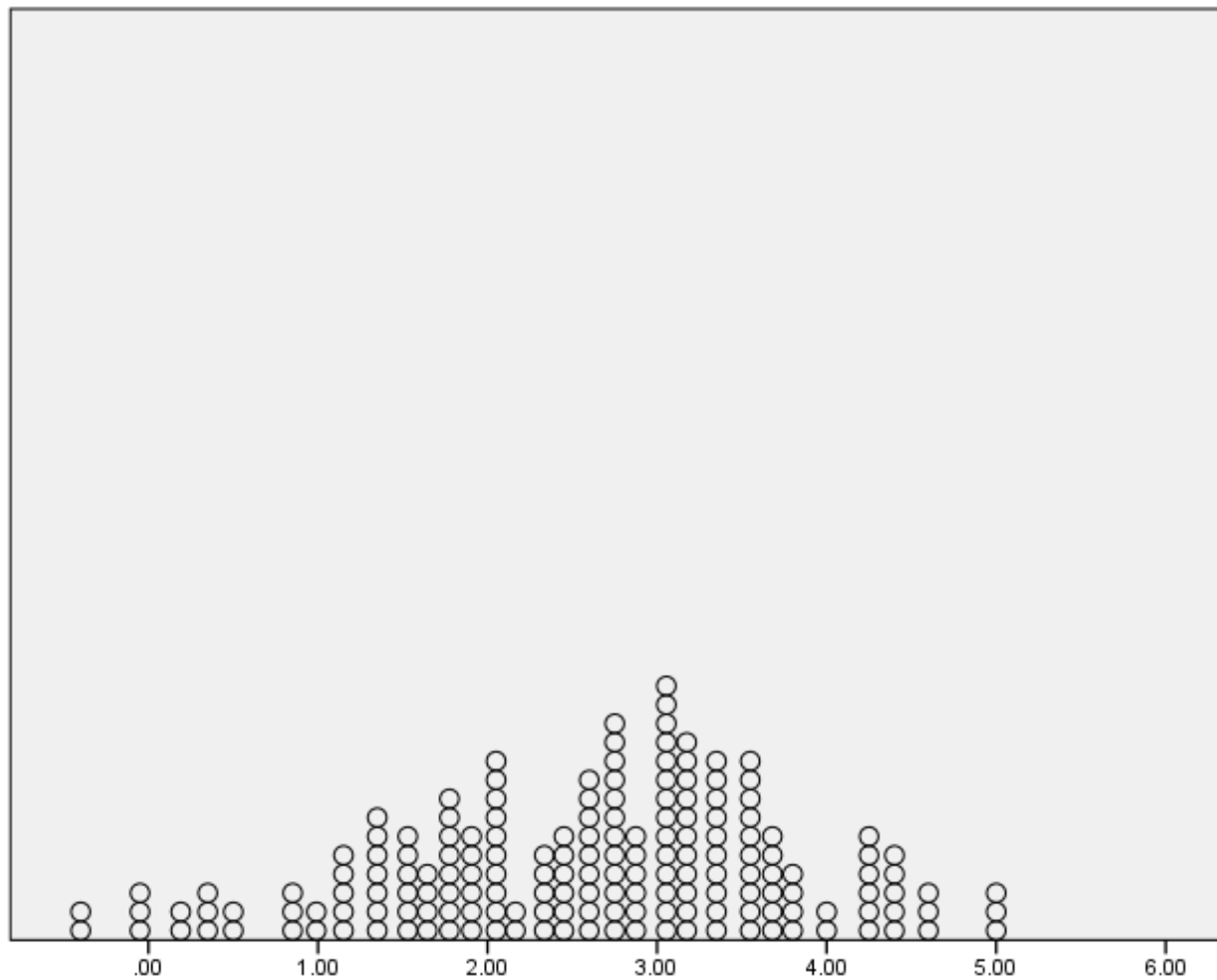


Figure 2. Difference between students' scores on comprehension in the fall and the average difficulty level of books interacted with in the fall each dot represents one student, and the scale on the X-axis is in grade-levels.

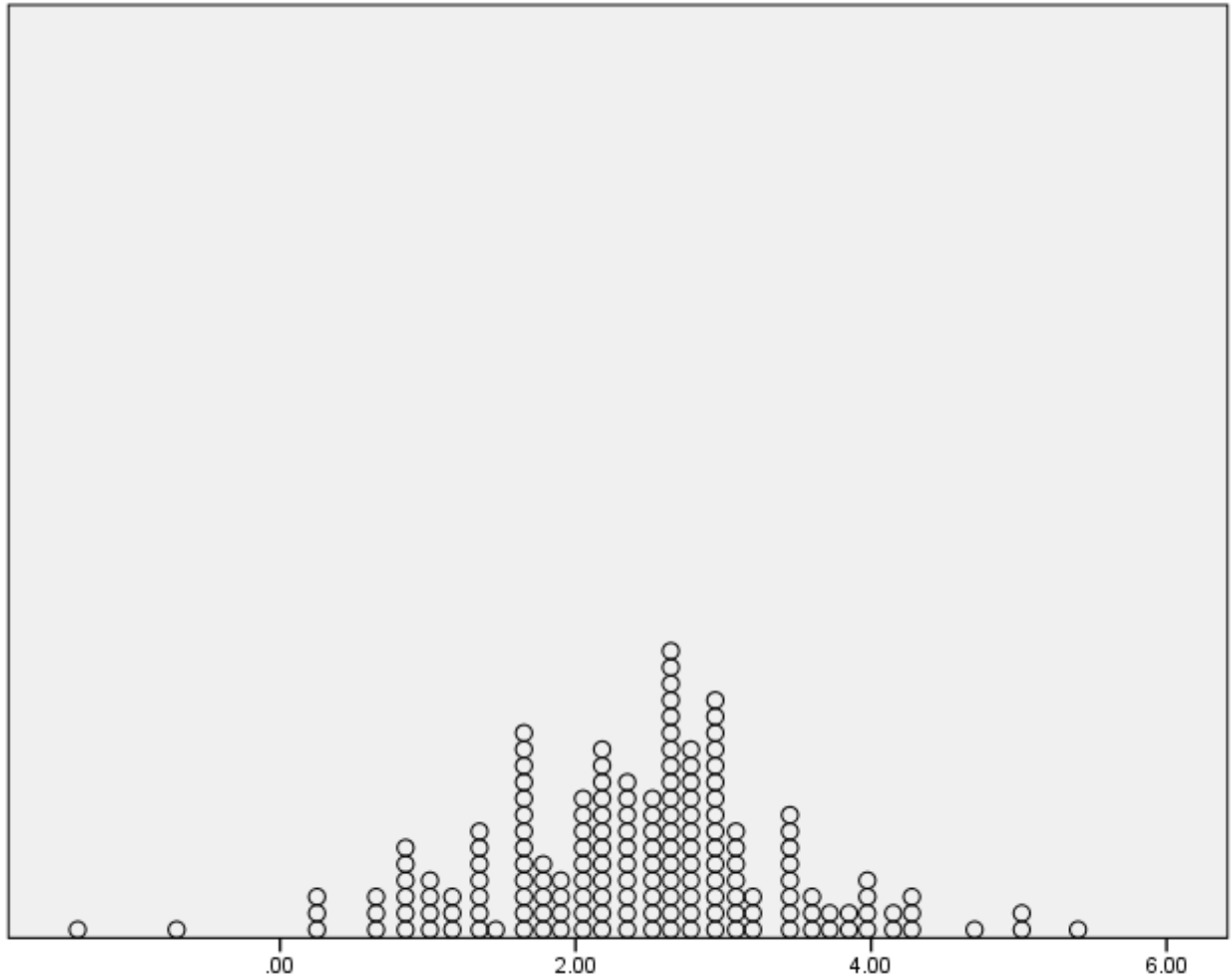


Figure 3. Difference between students' scores on decoding in the spring and the average difficulty level of books interacted with in the fall each dot represents one student, and the scale on the X-axis is in grade-levels.

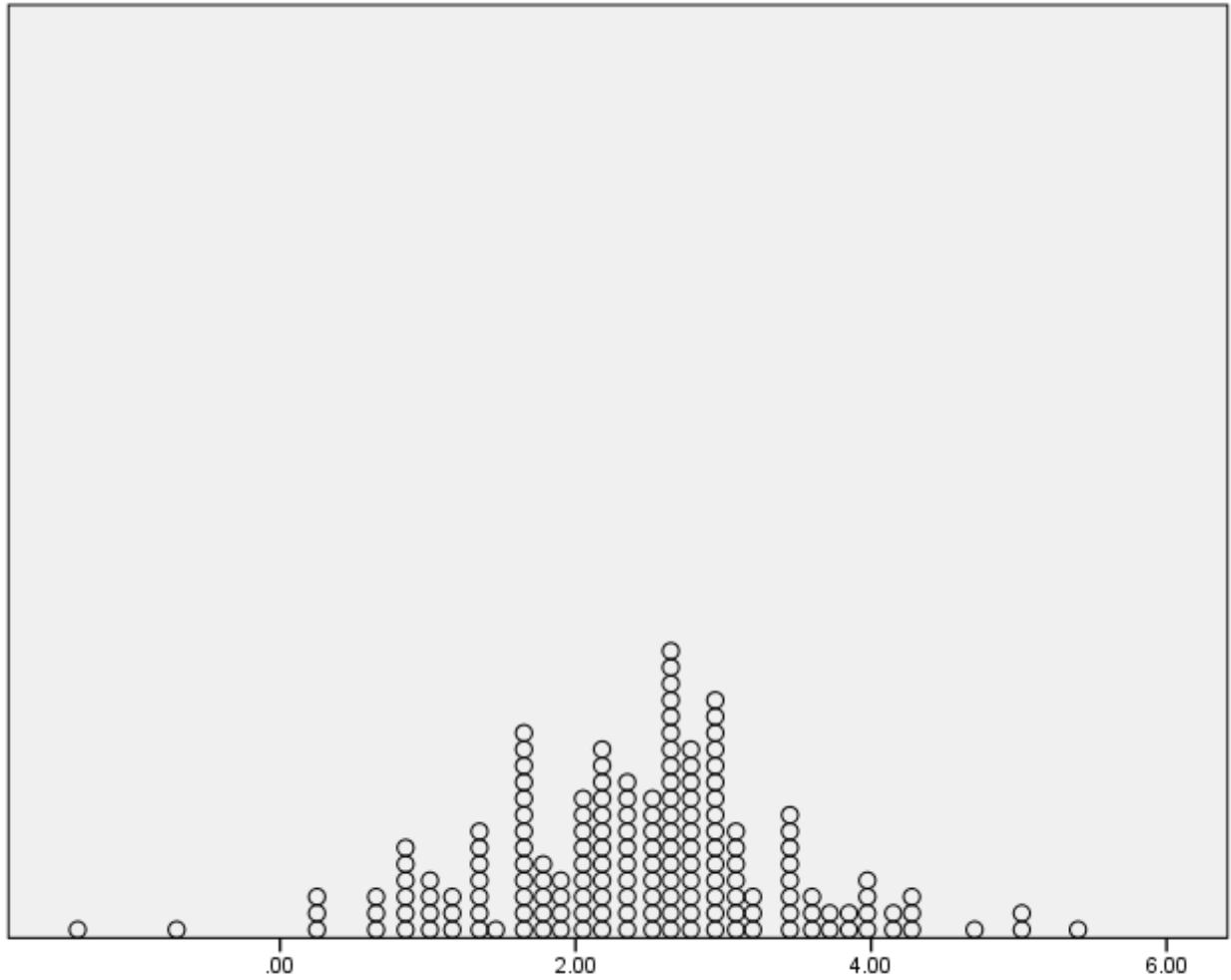


Figure 4. Difference between students' scores on comprehension in the spring and the average difficulty level of books interacted with in the fall each dot represents one student, and the scale on the X-axis is in grade-levels.

References

- 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.
- Battistich, V., Watson, M., Solomon, D., Lewis, C., & Schaps, E. (1999). Beyond the three R's: A broader agenda for school reform. *Elementary School Journal, 99*, 415–431.
- Begeny, J. C., Krouse, H. E., Brown, K. G., & Mann, C. M. (2011). Teacher judgments of students' reading abilities across a continuum of rating methods and achievement measures. *School Psychology Review, 40*(1), 23-38.
- Bisanz, J., Morrison, F. J., & Dunn, M. (1995). Effects of age and schooling on the acquisition of elementary quantitative skills. *Developmental Psychology, 31*, 221–236.
doi:10.1037/0012-1649.31.2.221
- Brenner, D., & Hiebert, E. H. (2010). If I follow the teachers' editions, isn't that enough? Analyzing reading volume in six core reading programs. *The Elementary School Journal, 110*(3), 358-363. doi:10.1086/648982
- Buttaro, A., Catsambis, S., Mulkey, L., & Steelman, L. C. (2010). An organizational perspective on the origins of instructional segregation: School composition and use of within-class ability grouping in American kindergartens. *Teachers College Record, 112*, 1300-1337.
- Christian, K., Morrison, F. J., Frazier, J. A., & Massetti, G. (2000). Specificity in the nature and timing of cognitive growth in kindergarten and first grade. *Journal of Cognition and Development, 1*, 429–448. doi:10.1207/S15327647JCD0104 04

- Connor, C. M., Morrison, F. J., Fishman, B. J., Ponitz, C. C., Glasney, S., Underwood, P. S., Piasata, S. B., Crow, E. D., & Schatschneider, C. (2009). The ISI Classroom Observation System: Examining the literacy instruction provided to individual students. *Educational Research, 38*(2), 85-99. doi:10.3102/0013189X09332373.
- Connor, C., Morrison, F., & Slominski, L. (2006). Preschool instruction and children's emergent literacy growth. *Journal of Educational Psychology, 98*(4), 665-689. doi:10.1037/0022-0663.98.4.665.
- Croninger, R.G., Rice, J.K., Rathbun, A., & Nishio, M. (2007). Teacher qualifications and early learning: Effects of certification, degree, and experience on first-grade student achievement. *Economics of Education Review, 26*, 312-324. doi:10.1016/j.econedurev.2005.05.008.
- Cunningham, A. E., Zibulsky, J., Stanovich, K. E., & Stanovich, P. J. (2009). How teachers would spend their time teaching language arts: The mismatch between self-reported and best practices. *Journal Of Learning Disabilities, 42*(5), 418-430. doi:10.1177/0022219409339063
- Cunningham, J. W., Spadorcia, S. A., Erickson, K. A., Koppenhaver, D. A., Sturm, J. M., & Yoder, D. E. (2005). Investigating the instructional supportiveness of leveled texts. *Reading Research Quarterly, 40*(4), 410-427. doi:10.1598/RRQ.40.4.2
- Darling-Hammond, L., Berry, B., & Thoreson, A. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis, 23*(1), 57-77.
- Doyle, W. (1977). Paradigms for research on teacher effectiveness. *Review of research in education, 5*, 163-198.

Early, D. M., Maxwell, K. L., Burchinal, M., Bender, R. H., Ebanks, C., Henry, G. T., & ... Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78*(2), 558-580. doi:10.1111/j.1467-8624.2007.01014.x

Flesch, R. (2006). 1948 - the flesch formulas. In W. DuBay (Ed.), *The classic readability studies*.

Goodlad, J. I. (1984). *A Place Called School. Prospects for the Future*. McGraw-Hill Book Company: New York, NY.

Guthrie, S. J., Connor, C. M., Morrison, F. J. (2013). *The cumulative impact of effective instruction from preschool through second grade*. Manuscript in revision.

Guthrie, S. J., & Morrison, F. J. (2012). *Exploring child by instruction interactions in kindergarten: A description of kindergarten instruction and child outcomes*. Manuscript in review.

Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy, 100*(1), 84-117.

reading achievement: The role of gender, classroom behavior, and emergent literacy skills in a longitudinal sample of children exposed to poverty. *Reading and Writing, 15*(7-8), 789-809. doi:10.1023/A:1020985701556

Hiebert, E. H., Martin, L., & Menon, S. (2005). Are There Alternatives in Reading Textbooks? An Examination of Three Beginning Reading Programs. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 21*(1), 7-32. doi:10.1080/10573560590523621

Huang, F. L., & Moon, T. R. (2009). Is experience the best teacher? A multilevel analysis of teacher characteristics and student achievement in low performing schools. *Educational*

Assessment, Evaluation and Accountability, 21(3), 209-234. doi:10.1007/s11092-009-9074-2

Hiebert, E. H., & Pearson, P. D. (2010). An examination of current text difficulty indices with early reading texts. *Reading Research Report*, 10.

Kane, T. J., & Staiger, D. O. (2012). Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains. Policy and Practice Brief. MET Project. *Bill & Melinda Gates Foundation*.

La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The Classroom Assessment Scoring System: Findings from the prekindergarten year. *The Elementary School Journal*, 104(5), 409-426. doi:10.1086/499760

Mashburn, A. J., Hamre, B. K., Downer, J. T., & Pianta, R. C. (2006). Teacher and Classroom Characteristics Associated With Teachers' Ratings of Prekindergartners' Relationships and Behaviors. *Journal of Psychoeducational Assessment*, 24(4), 367-380. doi:10.1177/0734282906290594

Mather, N., Woodcock, R. W. (2001). Examiner's Manual. *Woodcock-Johnson III Tests of Achievement*. Ithasca, IL:Riverside Publishing.

McNamara, D., Graesser, A., Cai, Z., Kulikowich, J., & McCarthy, P. (February 2010). Coh-Metrix (Report to the Gates Foundation/Text Complexity Project). Memphis, TN: University of Memphis.

Morrison, F. J., Cameron, C., Connor, C. M., Strasser, K., Griffin, E. (2005). Pathways to Literacy: Classroom Observation Coding Manual [Training videos and code book]. Unpublished instrument. University of Michigan: Pathways to Literacy Lab.

- Morrison, F. J., Griffith, E. M., & Alberts, D. M. (1997). Nature-nurture in the classroom: Entrance age, school readiness, and learning in children. *Developmental Psychology, 33*, 254–262. doi:10.1037/0012-1649.33.2.254
- Morrison, F. J., Griffith, E. M., & Frazier, J. A. (1996). Schooling and the 5 to 7 shift: A natural experiment. In A. J. Sameroff, & M. M. Haith (Eds.), *The five to seven year shift: The age of reason and responsibility* (2nd ed., pp. 161–186). Chicago, IL: University of Chicago Press.
- Morrison, F. J., Smith, L., & Dow-Ehrensberger, M. (1995). Education and cognitive development: A natural experiment. *Developmental Psychology, 31*, 789–799. doi:10.1037/0012-1649.31.5.789
- National Center for Education Statistics (2009). *Basic reading skills and the literacy of America's least literate adults: Results from the 2003 National Assessment of Adult Literacy Supplemental Studies*. Retrieved from nces.ed.gov/pubs2009/2009481.pdf
- National Institute for Child Health and Human Development, New York State Science and Technology Foundation, & Zeno, S. (1995). *Educator's word frequency guide*. Apple Valley, MN: Touchstone Applied Science Association.
- National Institute of Child Health and Human Development (NICHD), Early Child Care Research Network (ECCRN). (2005). A day in third grade: A large-scale study of classroom quality and teacher and student behavior. *The Elementary School Journal, 105*, 305-323.
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004) How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*, 237-257. doi:10.3102/01623737026003237.

- Peterson, P. L., & Fennema, E. (1985). Effective teaching, student engagement in classroom activities, and sex-related differences in learning mathematics. *American Educational Research Journal*, 22(3), 309-335.
- Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal*, 45(2), 365-397. doi:10.3102/0002831207308230.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity. *Educational Researcher*, 38: 109-119, doi:10.3102/0013189X09332374
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System™: Manual K-3*. Baltimore, MD US: Paul H Brookes Publishing.
- Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L. (2008). Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early childhood research quarterly*, 23(4), 431-451.
- Ponitz, C., & Rimm-Kaufman, S. E. (2011). Contexts of reading instruction: Implications for literacy skills and kindergarteners' behavioral engagement. *Early Childhood Research Quarterly*, 26(2). doi:10.1016/j.ecresq.2010.10.002
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (Second Ed.)*. Thousand Oaks, CA: Sage.
- Ryans, D. G. (1961). Some relationships between pupil behavior and certain teacher characteristics. *Journal of Educational Psychology*, 52(2), doi:10.1037/h0040990.
- Ryans, D. G. (1963). Assessment of teacher behavior and instruction. *Review of Educational Research*, 33(4), 415-441.

- Sadoski, M., & Willson, V. L. (2006). Effects of a Theoretically Based Large-Scale Reading Intervention in a Multicultural Urban School District. *American Educational Research Journal*, 43(1), 137-154. doi:10.3102/00028312043001137
- Sameroff, A. J., & MacKenzie, M. J. (2003). Research strategies for capturing transactional models of development: The limits of the possible. *Development and Psychopathology*, 15(3), 613-640. doi:10.1017/S0954579403000312
- Sanders, W. L., Wright, S. P., & Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of personnel evaluation in education*, 11(1), 57-67.
- Shepherd, M. A. (2011). Effects of ethnicity and gender on teachers' evaluation of students' spoken responses. *Urban Education*, 46(5), 1011-1028. doi:10.1177/0042085911400325
- Skibbe, L. E., Grimm, K. J., Stanton-Chapman, T. L., Justice, L. M., Pence, K. L., & Bowles, R. P. (2008). Reading trajectories of children with language difficulties from preschool through grade five. *Language, Speech, and Hearing Services in Schools*, 39, 475-486. doi:10.1044/0161-1461(2008/07-0016
- Sonnenschein, S., Stapleton, L. M., & Benson, A. (2010). The relation between the type and amount of instruction and growth in children's reading competencies. *American Educational Research Journal*, 47(2), 358-389. doi:10.3102/0002831209349215
- Stenner, A. J., Burdick, H., Sanford, E. E., & Burdick, D. S. (2007). The Lexile framework for reading (Technical report). Durham, NC: Metametrics.