2016

Classroom Organization by Prior Performance Interactions as Predictors of Literacy and Language Achievement

Heather Pilcher

FLORIDA STATE UNIVERSITY

COLLEGE OF EDUCATION

CLASSROOM ORGANIZATION BY PRIOR PERFORMANCE INTERACTIONS AS PREDICTORS OF LITERACY AND LANGUAGE ACHIEVEMENT

By

HEATHER PILCHER

A Dissertation submitted to the
Department of Teacher Education
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2016

Heather Pilcher defended this dissertation on March 25, 2016.

The members of the supervisory committee were:

Young-Suk Grace Kim

Professor Directing Dissertation

Carla Wood

University Representative

Barbara Foorman

Committee Member

Alysia Roehrig

Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with university requirements.

This work would not have been possible without the love and support of my wonderful network of family and friends. I would especially like to thank my wonderful husband, Rick Pilcher, for his unwavering support of me in all things.

ACKNOWLEDGMENTS

Thank you to the Institute for Education Science Pre-doctoral Interdisciplinary Research Training Program through the Florida Center for Reading Research for funding the first two years of my doctoral study. I would also like to acknowledge my major professor, Dr. Young-Suk Grace Kim, for allowing me to work as a part of research projects and for funding the last two years of my doctoral study. To all of the members of my committee, thank you for having high expectations for my work throughout this process and for having the patience to help me meet those expectations.

I would also like to thank my coworkers in Dr. Kim's lab for their efforts at collecting, coding, and otherwise taming the data that were used in this study. There are too many of you to name, but you are all appreciated.

Finally, I would like to thank the children, teachers, and district and school administrators who graciously participated in this research project. Thank you so much for allowing us to invade your space and steal your time.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Teachers' interactions with children represent an important source of influence in children's learning and development. Classroom organization, or the way the teacher manages the physical and behavioral aspects of the classroom environment, is one way that teachers can provide needed support to students who might otherwise struggle to be successful in the classroom environment. It is hypothesized to facilitate more time spent working productively on academic tasks which, in turn, is associated with higher academic achievement. In this study, teachers' classroom organization was observed and rated in the first, second, and third grades, and students' word reading, reading comprehension, and listening comprehension were measured each year. Cross-sectional analyses of the second and third grade data were conducted in order to replicate the findings of a previous study using the first grade data from the same data set (Pilcher & Kim, 2015). The relations of ratings of classroom organization and the interactions of classroom organization with children's prior performance (measured in the fall of first grade) in each year were also examined longitudinally. The sample included 264 students and 29 teachers in year one, 36 teachers in year two, and 35 teachers in year three. In the second and third grade cross-sectional analyses, classroom organization was not associated with student outcomes, and no significant interactions of classroom organization with prior achievement were detected. No significant longitudinal relations of classroom organization with student achievement were detected, but there were two significant interaction effects. Second grade teachers' classroom organization interacted with prior performance to predict students' third grade word reading and reading comprehension such that students whose scores were below the mean in the fall of the first grade had third grade scores that were very similar or only slightly higher when their second grade teachers provided higher levels of classroom organization. Students who began first grade

with word reading and/or reading comprehension scores that were above the mean had third grade scores that were much higher when their teachers provided high levels of classroom organization in the second grade. The most likely explanation for these findings seems to be that children who leave first grade without the prerequisite skills for engaging in meaning-based instruction in second grade are unable to benefit from having a second grade teacher who provides a high level of classroom organization.

CHAPTER ONE

INTRODUCTION

The primary grades are important years for children as they begin formal reading instruction. Growth trajectories in reading are established during these early grades, and children who are behind during the primary grades tend to have problems catching up to their peers in later years (Jimerson, Egeland, & Teo, 1999; Juel, 1988; McClelland, Acock, & Morrison, 2006). High quality teacher-child interactions have been hypothesized to serve a protective role for lower performing students because the interactions may compensate for home life experiences that are not available to some of these students (Jimerson et al., 1999). In fact, Sanders and Rivers (1996) stated that there may be residual effects of highly effective or ineffective teachers that can be detected as much as two years later. Therefore, high quality teacher-child interactions in the primary grades may provide struggling students with experiences that could help close the gap between them and their higher achieving peers at a key point in their development of reading proficiency.

Previous studies have found significant amounts of variation in the way that teachers manage their classrooms and interact with their students (e.g., Connor et al., 2014; Hamre et al., 2013; NICHD, 2002; Wright et al., 1997). In addition, classroom instruction has been shown to explain some of the variation in students' reading and language arts achievement (Cameron, Connor, Morrison, & Jewkes, 2008; Connor et al., 2014; Foorman et al., 2006; Mehta, Foorman, Branum-Martin, & Taylor, 2005). These findings are in line with the bioecological model of child development (Bronfenbrenner & Morris, 1998). The bioecological model states that children's achievement in school is strongly influenced by proximal factors (e.g., teacher-student interactions) and only weakly influenced by distal factors (e.g., district policies). Thus,

classroom and teacher characteristics have been shown to be significant proximal factors that are predictive of child outcomes (Cameron et al., 2008; Chatterji, 2006; Connor et al., 2014; Hamre & Pianta, 2005).

Emerging evidence from twin studies shows that teachers have a significant relation to student achievement beyond the influence of genetics (Taylor, Roehrig, Soden Hensler, Connor, & Schatschneider, 2010). In a value-added analysis, Taylor and colleagues found that when monozygotic twins were placed in different classrooms, twins whose oral reading fluency score was below the mean were in classrooms where teacher value-added scores were lower, and twins who scored above the mean were in classrooms where teacher value-added scores were higher. Therefore, the authors concluded that differences in teacher and classroom characteristics predicted differences in oral reading fluency scores within twin pairs and that differences between teachers have a measurable relation to their students' reading achievement over and above the influence of genetics.

Rutter and Maughan (2002) hypothesized that "gene-environment interactions mean that large individual differences are very likely to be found with respect to children's responses to specific teaching inputs and to the overall school environment" (p. 459). Interactions between child characteristics and teacher or classroom characteristics have been documented in the teacher quality literature (e.g., Cadima, Leal, & Burchinal, 2010; Cameron et al., 2005; Connor et al., 2014; Hamre & Pianta, 2005). The literature is rich with descriptions of high quality teaching practices (e.g., see reviews presented in Bohn, Roehrig, & Pressley, 2004; Stronge, Ward, & Grant, 2011; Taylor & Pearson, 2002), but more information is needed about the complex interactions of child characteristics and teacher or classroom characteristics.

Studies of the quality of teaching practices are mainly cross-sectional in nature (e.g., Hamre et al., 2012; Li Grining et al., 2010; Mashburn et al., 2006; Pianta et al., 2005), and only a few studies have followed the same group of students for multiple years (Cadima et al., 2010; Curby, Rimm-Kaufman, & Ponitz, 2009; Hamre & Pianta, 2001). Some studies have included observations of second or third grade classrooms, but have focused mainly on one sub-group of students, such as students from high-poverty schools (Foorman et al., 2006) or have included observations without linking them to student achievement (NICHD, 2005).

Furthermore, much work on the quality of teaching practices has been primarily with young children such as those in preschool (Hamre & Pianta, 2001; Hamre et al., 2012; Li Grining et al., 2010; Mashburn et al., 2006; Pianta et al., 2005), kindergarten (Curby, Rimm-Kaufman, & Ponitz, 2009; Hamre & Pianta, 2001; La Paro et al., 2009), and first grade (Cameron, Connor, Morrison, & Jewkes, 2008; Curby et al., 2009; Hamre & Pianta, 2001; Hamre & Pianta, 2005; NICHD, 2002). Less is known about second grade and third grade.

The current longitudinal study included observations of the instruction that students experienced in the first, second, and third grades and examined the relations between the quality of classroom organization and student achievement as well as the interactions of students' prior performance and the quality of classroom organization that they experienced. A previous study using only the first grade data from the same dataset yielded promising results for ratings of classroom organization as a predictor of student achievement. Pilcher and Kim (2015) found that ratings derived from the classroom organization rating scale were significantly and positively related to students' reading and listening comprehension at the end of the first grade. Classroom organization ratings also interacted with children's fall word reading scores such that children in the 25th percentile had spring scores that were most like their peers when their teachers provided

high levels of classroom organization. The current study was conducted in order to extend these findings longitudinally from first through third grade.

CHAPTER TWO

LITERATURE REVIEW

Interest in research on teacher effectiveness peaked in the 1980's and 1990's when a large body of literature was published containing correlational and descriptive studies of effective teaching practices. Since then, this body of research has shown that there are some common characteristics among teachers who have proven to be effective at increasing students' literacy achievement (Wang, Haertel, & Walberg, 1993). These characteristics may be non-academic in nature and are not necessarily related to teachers' background characteristics, such as years of teaching experience (Stronge, Ward, & Grant, 2011). Teachers who are in the top quartile of effectiveness are described as having higher levels of classroom management and organization and engaging in more positive interactions with their students (Stronge et al., 2011).

One common theme in the literature is that effective teachers show high levels of responsiveness to the needs of their students while using effective behavior management techniques (Hamre et al., 2013). Positive academic and social interactions between teachers and students have been linked to positive student learning outcomes. Teachers who model and encourage higher levels of fairness and respect and also have positive relationships with their students are associated with high reading achievement (Stronge et al., 2011). These teachers have been described as having consistent expectations and consequences for misbehavior, redirecting problem behaviors positively, and having fewer behavior problems in their classes (Stronge et al., 2011; Wharton-McDonald, Pressley, & Hampston, 1998), since behavior problems are less frequent when the teacher uses classroom management effectively.

A second common characteristic of effective teachers is that they are masters of classroom organization. Classroom organization refers to "how order is established and

maintained in classroom environments" (Doyle, 1986). In previous studies of high quality instruction, effective teachers were found to be better managers and organizers, encouraged students to have more responsibility in the classroom, were consistently well-prepared, had predictable daily classroom routines, were able to transition between activities quickly, handled routine tasks promptly and efficiently, and prevented or minimized interruptions (Wang et al., 1993; Stronge et al., 2011; Wharton-McDonald et al., 1998). The literature also suggests that classrooms in which students hold some responsibility for the management of the classroom routines are more effective (Doyle, 1986). Lessons run more smoothly when the necessary materials are readily available for student use and when the physical layout of the room is appropriate (Stronge et al., 2011).

Classroom management involves promoting order in the classroom rather than simply reacting to misbehavior or low student engagement when it occurs (Doyle, 1986). When a classroom is well managed from the beginning of the school year, misbehavior is usually rare, and teachers who handle misbehavior most effectively are more likely to have students who make learning gains. These teachers keep the entire class involved in the lesson, through effective management, so that misbehavior is minimized (Doyle, 1986) and they are less likely to show anger or irritation when correcting misbehavior (Anderson, Evertson, & Brophy, 1979).

Teacher Observation Instruments

The Measure of Effective Teaching, or MET project (Kane & Staiger, 2012), is probably the largest study of teacher observation instruments to date that has linked observations of teacher behavior to students' academic achievement outcomes. Teachers in a variety of grade levels and classroom settings were observed multiple times and rated using several different teacher observation instruments. Two instruments were related to general instructional practices

in the primary grades and beyond: Framework for Teaching (FFT; Danielson, 2013) and Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008). Both of the instruments were found to be positively related to student achievement gains and the reliability of the observation scores was found to be acceptable when multiple observations were used.

The area of interest in the current study, classroom and behavior management, was generally rated as average to high quality in the MET study (Kane & Staiger, 2012) for both of the instruments mentioned above. Factor analyses of each instrument showed that all instruments had a similar factor structure. The first factor was a general factor that seemed to represent general teaching competency. The second factor represented classroom and behavior management and the third factor represented the unique aspects of each instrument. These analyses provide evidence that classroom and behavior management is a factor that can be observed separately from other competencies in a teacher observation instrument.

Danielson's Framework for Teaching (FFT; Danielson, 2013) is used in many school districts in the United States as well as in a few other countries worldwide. It includes four domains (planning and preparation, classroom environment, instruction, and professional responsibilities) which each include five to six components on which teachers can be scored from level 1 (unsatisfactory) to level 4 (distinguished). Domains two (classroom environment) and three (instruction) are most closely aligned with the content of the classroom organization rating scale used in the current study. For example, the indicator "managing classroom procedures" within domain two has the items "management of transitions" and "management of materials and supplies," which are aligned with the current study items "there is a system in place for facilitating students' transition from one part of the room to another" and "students

seem to be following a set of procedures for using classroom materials that have previously been taught," respectively.

The Marzano framework (Marzano, 2007) is another teacher evaluation instrument currently in use in many school districts. The framework is made up of ten instructional design questions and each question is based on a collection of meta-analyses and research syntheses. Several of Marzano's (2007) instructional design questions are similar in content to the items used in the current study. For example, the design question "What will I do to engage students?" includes behaviors such as managing questions and response rates, using appropriate pacing, and providing opportunities for students to talk about themselves. These behaviors all align with one or more items on the scale used in the current study. Other design questions from the Marzano framework that align with content on the scale used in the current study include: "What will I do to establish or maintain classroom rules and procedures?", "What will I do to recognize and acknowledge adherence and lack of adherence to classroom rules and procedures?", and "What will I do to establish and maintain effective relationships with students?"

The CLASS (Classroom Observation and Scoring System) teacher observation instrument is used most often in research contexts and is organized according to the "teaching through interactions" framework (Hamre et al., 2013). The teaching through interactions framework posits that the quality of teacher-student interactions is predictive of children's academic and social outcomes throughout schooling (Downer, Sabol, & Hamre, 2010; Hamre et al., 2013). The teaching through interactions framework has been divided into three aspects: emotional support, classroom organization, and instructional support (Hamre et al., 2013) and the three aspects have been shown to be malleable traits in intervention studies (Brown, Jones, LaRusso, & Aber, 2010; Curby, Brock, & Hamre, 2013; Hamre et al., 2012).

The three-factor model of the teaching through interactions framework has been supported in the literature (Hamre & Pianta, 2007; Hamre et al., 2013) and has been shown to have a better fit to the data than a one or two factor model (Hamre et al., 2013). However, even with the three factor model, the model-fit indices are not ideal in most cases (Hamre et al., 2013; Pakarinen et al., 2010) and the three factors tend to be highly correlated with each other (Rudasill, Gallagher, & White, 2010). Recent work has been done using the CLASS measure (Hamre, Hatfield, Pianta, & Jamil, 2014) in an attempt to resolve these issues. A bifactor model using CLASS data indicated that there may be a general factor that represents a teachers' general responsiveness to the children in their classrooms and other domain-specific factors that represent behaviors that are independent from the general factor (Hamre et al., 2014). Hamre et al. (2014) found that the best fitting model for CLASS data from a sample of preschool classes was the bifactor model with one general domain-specific factor representing the responsiveness of the teacher and two domain-specific factors (i.e., positive management and routines; cognitive facilitation). This model had superior fit compared to one-factor, two-factor, and three-factor confirmatory factor analyses. The general responsiveness factor was associated with more literacy and language skill growth, improved working memory, and a decrease in teacher-child conflict. The positive management and routines factor was associated with gains in inhibitory control and the cognitive facilitation factor was associated with early literacy and language skills. These findings are in agreement with the factor analyses described above that were part of the MET project (Kane & Staiger, 2012).

The classroom organization rating scale used in this study (see Appendix A) was adapted from a holistic rating scale used in a study of literacy instruction in third grade (Connor et al., 2014) that was similar in design to the current study. Connor and colleagues created the rating

scale based on a review of existing studies of effective classroom instruction. Holistic quality ratings have been found to capture stable features of classroom environments (Hamre & Pianta, 2005), especially when multiple observations and/or raters are used (Kane & Staiger, 2012). Connor and colleagues (2014) found that the quality ratings had a marginally significant relation to third graders' vocabulary and comprehension outcomes and there were several significant interactions of quality ratings with the amount of time that individual children spent on various types of literacy instruction. A more detailed description of the quality rating scale adaptations is provided in the Methods section.

The adapted version of the scale was used by Pilcher and Kim (2015) in a study of first grade teachers and students. Classroom organization factor scores derived from the scale had significant and positive relations to first graders' reading and listening comprehension at the end of first grade. A significant interaction was found between students' fall word reading scores and classroom organization that was predictive of students' spring word reading scores. Lower performing students had word reading scores that were the most similar to those of their higher performing peers when their teachers provided higher levels of classroom organization. The current study extended the findings from Pilcher and Kim (2015) by including longitudinal data from the first through third grades.

The classroom organization rating scale most likely captures aspects of the general responsiveness domain described by Hamre et al. (2014) as well as having some overlap with the positive management and routines and the cognitive facilitation domains. For example, the item "the pacing of the activities is appropriate for the students" is very similar to one aspect of what Hamre and colleagues (2014) call the general responsiveness factor. The item "procedures are evident for daily classroom tasks" is similar to one aspect of what the authors describe as a

domain-specific factor called "positive management and routines." There are even a few items that seem to overlap somewhat with the domain-specific factor called "cognitive facilitation." For example, "the teacher incorporates students' examples and experiences into the lesson" relates very closely to Hamre et al.'s (2014) description of concept development, which is one aspect of the cognitive facilitation factor. In the current study, the items on the classroom organization rating scale are anticipated to be related to one factor which will be most similar to the general responsiveness factor with some overlap happening with the other two domain-specific factors described by Hamre et al. (2014).

Classroom Organization

The goal of the current study was to focus on the classroom organization domain and to examine its longitudinal relations to children's literacy and language achievement in the primary grades. The classroom organization domain was defined as a set of teacher behaviors that are supportive of positive student behavior during classroom interactions and are meant to result in more time spent productively on academic instruction (Hamre et al., 2013, Connor et al., 2014; Foorman et al., 2006).

Foorman et al. (2006) found that teachers who received lower global ratings of teacher quality (with organization being one aspect that was considered) were more likely to engage in behaviors that resulted in lost instructional time. These teachers were more likely to engage in what the authors coded as non-reading activities (e.g., transitions, discipline, grading papers). This finding provides support for the hypothesis in the current study that teachers who effectively manage the organizational aspects of the classroom will have higher performing students because they are able to use their instructional time more efficiently. Thus, higher levels of classroom organization are hypothesized to relate to student achievement because increased

amounts of time spent on academic instruction are associated with higher academic achievement, particularly when aspects of high quality teacher-student interactions (such as classroom organization) are taken into account (Connor, Morrison, & Katch, 2004; Connor, Morrison, & Underwood, 2007; Connor et al., 2014; Foorman et al., 2006).

There is some evidence that lower performing students in particular are able to spend more time working productively on academic tasks when their teachers provide more organization throughout the school year (Cameron et al., 2005).Cameron and colleagues (2008) found that when teachers focused on organization at the beginning of the year followed by sharply decreasing the amount of time spent on non-instructional activities for the rest of the year, first grade students ended the year with higher letter and word reading skills. They speculated that this positive relation was due to increased amount of time spent on reading and language arts instruction in highly organized classrooms.

High levels of classroom organization may also help make the time spent on instructional activities more beneficial minute-for-minute (Connor et al., 2014; Foorman et al., 2006). Foorman and colleagues (2006) found that more when teachers with higher ratings of teacher effectiveness spent more time on emergent literacy instruction first grade students had better word attack skills. In addition, Connor and colleagues (2014) found that increases time spent in teacher-managed whole class or small group meaning based instruction (vocabulary, reading comprehension, etc.) were associated with higher vocabulary and reading comprehension achievement for third grade students when their teachers provided higher levels of classroom organization.

In the current study, the label “classroom organization” was selected to represent the set of teacher behaviors and classroom characteristics that were rated. This label was intended to

reflect both the physical organization and management of the classroom as well as the way that classroom interactions are organized in order to create a positive social environment in which students interact with competent teachers. In the literature, labels such as "teacher quality" (Taylor et al., 2010), "teacher effectiveness" (Foorman et al., 2006), and "quality of teacher-student interactions" (Cadima, Leal, & Burchinal, 2010; Curby, Rimm-Kaufman, & Ponitz, 2009) have been used to describe more global observations about the teacher's competence. These labels were not selected in the current study because they lacked specificity about what exactly was being observed. In contrast, labels such as "positive management and routines" (Hamre et al., 2014) and "teacher organization" (Cameron, Connor, & Morrison, 2005) seem to be too specific to capture the range of behaviors that were observed in the current study.

Interactions of Prior Performance with Classroom Organization

Children's prior literacy performance is a strong predictor of their later literacy performance, and children with lower prior performance who experience high levels of classroom organization may have higher academic achievement than lower performing students who do not experience high levels of classroom organization (Cadima, Leal, & Burchinal, 2010; Cameron et al., 2005; Curby et al., 2009). So, beginning skill level predicts end of year outcomes differently depending on the level of classroom organization that the child experiences, that is, the child's characteristic (prior performance) interacts with a characteristic of the classroom instruction (classroom organization). Child by classroom instruction interactions have been observed in the literature on learning environment quality as well (Cadima, Leal, & Burchinal, 2010; Cameron et al., 2005; Curby et al., 2009).

Cameron et al. (2005) examined child by classroom instruction interactions in a sample of 44 first grade classrooms. Classroom organization was coded as the amount of time the

teacher spent organizing and orienting the students to classroom activities and the amount of time spent on classroom organization was found to decrease over the course of the school year. Students with higher initial vocabulary levels participated in more child-managed instruction when more time was spent on organization in the fall and then the amount of time spent on organization decreased sharply in the winter and spring. On the other hand, children with lower initial vocabulary scores needed more time spent on classroom organization. In other words, classroom organization in the winter interacted with children's beginning vocabulary levels such that children with lower initial vocabulary scores were able to participate in more child-managed instruction (time the child spends managing his or her own attention as opposed to needing support from the teacher) when classroom organization was high. So, children with low prior performance in vocabulary who were in classrooms where more time was spent on organization in the winter needed less support from the teacher during independent work in the spring than children with low prior performance in vocabulary who were in classrooms where less time was spent on organization in the winter. Cameron et al. (2005) suggested that children with lower prior performance levels need more classroom organization throughout the year and children with higher prior performance levels are more able to work independently with less support from the teacher.

This study provided evidence that children with different characteristics may respond to teachers' classroom practices differently. One weakness in this study is that the students' academic skills were not included as outcome measures, so the relation between organization and student achievement was not examined. While interactions were found to predict the amount of time students spent working independently, their relation with student achievement was also not examined. In the current study, relations between classroom organization and student

achievement will be examined as well as relations between interactions of classroom organization and prior performance with student achievement. Another weakness is that vocabulary was the only aspect of prior performance to be considered. An important question to consider is whether other academic skills (e.g., word reading) also have some relation to and interaction with classroom organization.

In another study of first grade classrooms, Cadima, Leal, and Burchinal (2010) examined the relations between prior performance in preschool, classroom organization (as measured by the CLASS), and first grade reading and math outcomes for 106 first grade students in Portugal. First grade classroom organization was a significant predictor of vocabulary and print concepts at the end of first grade, but the interaction of prior performance and classroom organization was not significant for any of the reading outcomes. The interaction of classroom organization and prior performance was predictive of the first grade math outcome such that children with lower prior performance in preschool had higher math scores at the end of first grade when classroom organization in first grade was high. These findings provide support for the hypothesis that the effects of classroom organization on academic achievement may be moderated by prior performance.

One reason that Cadima et al. (2010) may have failed to find significant interaction effects for literacy related measures was that the first grade classrooms were only observed once. Recent work on the reliability of teacher observations suggests that averaged scores from multiple classroom observations are much more reliable than single observations (Kane & Staiger, 2012). Finally, the students were sampled in such a way that each classroom cluster only included one to four students. Designs in which most clusters (classrooms) have a very small

number of participants may not have sufficient power to detect interaction effects (Snijders, 2005).

Curby et al. (2009) tested for child by instruction interactions in a longitudinal study across the kindergarten and first grade years. The sample included 147 children from mainly low income families. Thirty-six classrooms with four to seven children per classroom were included in the final study sample. The authors found several interactions of prior performance with aspects of high quality teacher-child interactions (as measured by the CLASS) that were predictive of first grade achievement. An interaction of prior performance in math with kindergarten classroom organization was predictive of first grade math achievement such that children with low prior performance levels in math were the most successful in first grade when kindergarten classroom organization was higher and children with high prior performance levels in math were the most successful in first grade when kindergarten classroom organization was lower. Taken together, there is evidence that prior performance may moderate the association between classroom organization and academic achievement, and the effects of prior performance by classroom organization interactions may be detectable as much as a year later.

A limitation in the Curby et al. (2009) study was that the small number of students in the final sample (n = 333 consented and n = 171 in the final sample) may have reduced the power to detect significant interaction effects for the literacy outcome measures (letter-word identification and sound awareness). The authors noted that the three dimensions of the CLASS (Pianta et al., 2008) measure are often strongly correlated with each other. Thus, multicollinearity could be a problem in studies that include all three dimensions as separate predictors. As mentioned above there is some debate about the underlying factor structure of the CLASS measure and a single factor model of overall competence or quality may be more appropriate (Hamre et al., 2014).

These studies (Cadima, Leal, & Burchinal, 2010; Cameron et al., 2005; Curby et al., 2009; Pilcher & Kim, 2015) provided evidence that interactions of prior performance and classroom organization are present in early childhood settings, but they did not examine the longitudinal relations of these interactions to student achievement. Only two studies (Cadima et al., 2010; Curby et al., 2009) were found that both included interactions of prior performance with classroom organization and followed a group of students for more than one school year. Longitudinal studies are needed in this body of literature in order to document how these interactions may change over time as children progress through the elementary grades. The long-term relations of prior performance, classroom organization, and language and literacy outcome measures have not been well documented in the existing literature.

Current Study

Interactions between prior performance and classroom organization in the first, second, and third grades were included in the current study in order to add to the above mentioned literature on child by instruction interactions. The cross-sectional analyses described below are an attempt to replicate Pilcher and Kim's (2015) findings with the second and third grade data. The longitudinal analyses described below examined whether classroom organization at each grade level interacted with children's first grade achievement to predict their literacy and language achievement at the end of third grade. So, child by instruction interactions in first, second, and third grade were examined longitudinally to test the hypotheses that children with lower prior performance levels would benefit more from highly organized classroom instruction than their more highly skilled peers and that the effects of these interactions in the first, second, and third grades would be detectable at the end of the third grade.

In the current study the classroom organization ratings were averaged across two observations, which should correspond to an increase in the reliability of the scores (Kane & Staiger, 2012). The current study also included more classrooms with 5 or more participating students each (93% in year 1, 75% in year 2, and 74% in year 3). An increase in the number of participants per cluster is one way to increase power in a multilevel analysis (Snijders, 2005).

This analysis adds to the current literature by extending the range of grade levels that are examined as well as following the same group of children over time. Interactions of prior performance with classroom organization have been found to predict math achievement and time spent in child-managed instruction, but the studies described above did not find significant interactions that predicted literacy or language achievement. The main goal of the current study was to add to the literature on child by classroom instruction interactions by providing evidence for prior performance by classroom organization interactions that are predictive of literacy and language achievement.

Word reading, reading comprehension, and listening comprehension were used as literacy and language outcome measures for the current study. The simple view of reading states that reading comprehension, which is the ultimate goal of reading, is the product of accurate word reading and listening comprehension (Catts, Adlof, & Ellis Weismer, 2006; Hoover & Gough, 1990; Joshi, Tao, Aaron, & Quiroz, 2012; Kim, 2015). Since reading comprehension, word reading, and listening comprehension are in the process of being developed in the first through third grade years, all three constructs were included in the current study in order to provide the most accurate representation of children's reading and language growth during this key point in their development. The addition of listening comprehension as an outcome measure is a new contribution to the literature in this area.

The specific research questions are:

1. Is classroom organization in grades 2 and 3 related to students' language and literacy skills?
2. Does the classroom organization provided by first, second, and third grade teachers have a statistically significant relation to students' literacy and language achievement levels at the end of the third grade?
3. Do the interactions of first, second, and third grade teachers' classroom organization with prior literacy and language performance predict students' achievement levels at the end of third grade?

For research question 1 I hypothesized that the findings in grades two and three would replicate those in Pilcher and Kim (2015). My hypothesis for research question 2 was that classroom organization would be (possibly weakly) positively associated with literacy and language outcomes at the end of third grade. Small effect sizes for studies of teacher-student interactions are commonplace in this body of literature (e.g., Cadima et al., 2010; Mashburn et al., 2008). For research question 3 I hypothesized that prior performance would interact with classroom organization such that students with lower prior performance would benefit more from higher levels of classroom organization than students with higher prior performance.

CHAPTER THREE

METHODS

In each of the six participating schools consent forms were sent home with all of the first grade students. All students who returned completed consent forms were allowed to participate in the study. Students also gave their assent to be tested each time they were pulled from their classroom. Teachers completed informed consent forms at the beginning of each school year and gave their assent to be videotaped each time an observation was conducted in their classroom. The current study was conducted as a secondary data analysis and IRB approval was obtained accordingly with the author as the principle investigator. The IRB approval letter can be found in Appendix B.

Student Participants

The data for this study were from a larger study (Kim & Petscher, in press) in which students were assessed on their reading and language skills during the first through third grades and video observations were conducted with their teachers twice per year (fall and spring; a total of 6 waves of data). The students were selected using convenience sampling from six schools in two school districts in the southeastern United States.

Table 1 contains descriptive demographic information about the students in the original study. In years two and three only students who had not been retained in the previous year were included in the study (n = 29 retained in first or second grade). So, in year two only second grade classrooms were observed and in year three only third grade classrooms were observed. There were 351 participating students in wave 1 of data collection, and the study concluded with 303 participating students at the end of wave 6. Of the students who remained in the study 25 were a year or more older than their peers (age 10 years or older at the end of third grade). In year 3, the

sample was approximately 50% male and approximately 60% were Caucasian, 25% were African-Americans and the remainder of the sample represented children who were Hispanic (~ 6%), Asian (~ 3%), or reported other or multiple race/ethnicity categories (~6%). In wave 1 about 10% of the sample received services for exceptional students (ESE) and about 51% of the sample received free or reduced price lunch (FRL). In wave 6 about 19% were receiving ESE services and about 47% were receiving FRL. English-language learners represented only about 1% of the participants.

The final sample used for analysis in the current study (n = 268) was comparable to the original sample on all of the demographic variables except age at the end of year 3. When students who were retained in first or second grade were removed, only 8 students remained who were a year or more older than their peers. The final sample was approximately 49% male and approximately 62% were Caucasian, 24% were African-Americans and the remainder of the sample represented children who were Hispanic (~ 6%), Asian (~ 3%), or reported other or multiple race/ethnicity categories (~5%). About 22% received ESE services and about 42% received FRL. English-language learners represented only about 1% of the participants. The average class size for the final sample was 9 students (SD = 3.40) in year 1, 7 students (SD = 3.65) in year 2, and 8 students (SD = 3.54) in year 3.

Literacy and Language Measures

Students were assessed twice per year using a battery of standardized literacy and language measures. The outcomes used in this study included measures of word reading, reading comprehension, and listening comprehension. Cronbach's alpha values are found in Table 2. Word reading was assessed using the WIAT-III Word Reading subtest and the Woodcock Johnson-III Letter-Word Identification subtest. The WIAT-III Word Reading subtest (Wechsler,

2009) required students to read words that are listed in order of difficulty. Students continued reading until they had misread four consecutive words. Cronbach's alphas remained constant at 0.95 over the course of data collection. The Woodcock Johnson-III Letter-Word Identification subtest (Woodcock, McGrew, & Mather, 2001) required students to identify letters and then words that increased in difficulty as the test progressed. Students continued reading until they had misread the last six consecutive items on a page. Cronbach's alphas ranged from 0.91 to 0.92.

Listening comprehension was assessed using the OWLS-II Listening Comprehension subtest and the Woodcock Johnson-III Oral Comprehension subtest. In the OWLS-II Listening Comprehension subtest (Carrow-Woolfolk, 2011) students listened to phrases or sentences that were presented orally and pointed to the picture that best matched the prompt. The assessment was discontinued when the student had missed four consecutive items. Cronbach's alphas ranged from 0.91 to 0.93. The Woodcock Johnson-III Oral Comprehension subtest (Woodcock, McGrew, & Mather, 2001) required students to listen to a spoken sentence and supply a missing word. Students continued until they had responded incorrectly to six consecutive items. Cronbach's alphas ranged from 0.74 to 0.79.

Reading comprehension was assessed using the WIAT-III Reading Comprehension subtest and the Woodcock Johnson-III Passage Comprehension subtest. For the WIAT-III Reading Comprehension subtest (Wechsler, 2009), students read short passages silently or aloud and then answered comprehension questions about them. In this study, students read three grade-appropriate stories during each assessment wave and answered all of the associated comprehension questions for those stories. Cronbach's alphas ranged from 0.83 to 0.88 across the six time points. The Woodcock Johnson-III Passage Comprehension subtest (Woodcock,

McGrew, & Mather, 2001) required students to read sentences and paragraphs and supply a missing word. Testing was discontinued when the student missed six consecutive items ending on a completed page. Cronbach's alphas ranged from 0.75 to 0.88.

Teacher Participants and Video Observations

The students' teachers were observed once in the fall and once in the spring. The sample of teachers included 29 first grade teachers, 38 second grade teachers, and 37 third grade teachers. Descriptive information about teachers (Table 3) was collected using a digital survey that was emailed to them early in the spring of each year. The first grade teachers had an average of about 16 years (*SD* = 13) of teaching experience with an average of 11 years (*SD* = 11) of experience in first grade. Approximately 43% of them reported having completed a graduate level degree, approximately 79% reported being certified in either Elementary Education (K-6) or Early Childhood Education (PK-3), and approximately 71% reported holding other endorsements or certificates. The second grade teachers had an average of about 13 years (*SD* = 11) of teaching experience with an average of 5 years (*SD* = 6) of experience in second grade. Approximately 21% of them reported having completed a graduate level degree, approximately 74% reported being certified in either Elementary Education (K-6) or Early Childhood Education (PK-3), and approximately 53% reported holding other endorsements or certificates. The third grade teachers had an average of about 12 years (*SD* = 8) of teaching experience with an average of 7 years (*SD* = 6) of experience in third grade. Approximately 30% of them reported having completed a graduate level degree, approximately 84% reported being certified in either Elementary Education (K-6) or Early Childhood Education (PK-3), and approximately 65% reported holding other endorsements or certificates. The other endorsements or certificates that teachers reported were: Reading, English for Speakers of Other Languages (ESOL), School

Social Work, Exceptional Student Education (ESE), Middle Grades Integrated, Gifted, and National Board Certification.

All of the teachers in the study used a commercially published reading and language arts curriculum. In year one of the study, five of the participating schools used the *SRA Imagine It!* (SRA/McGraw-Hill, 2007) reading series and one school used *Story Town* (Houghton Mifflin Harcourt, 2008); in year two of the study, three schools used *Reading Wonders* (McGraw-Hill, 2014), one school used *Journey's Common Core* (Houghton Mifflin Harcourt, 2014), one school used *Reading Street Common Core* (Pearson, 2013), and one school used *SRA Imagine It!* (SRA/McGraw-Hill, 2007); in year three of the study, four schools used *Reading Wonders* (McGraw-Hill, 2014), one school used *Journey's Common Core* (Houghton Mifflin Harcourt, 2014), and one school used *Reading Street Common Core* (Pearson, 2013).

All videos were collected using two camcorders with wide angle lenses. Videographers received two hours of training and shadowed other videographers at least once before being sent out to videotape on their own. They were trained to take notes about the classroom activities, the physical descriptions of the students, and the layout of the classroom. Digital cameras were used to take photos of samples of students' work, their textbook pages, and center/small group activities. Videographers received detailed feedback on their performance after their first time videotaping on their own and then received periodic spot check feedback after that. Reminders and suggestions for improvement were discussed with videographers prior to the beginning of each wave of data collection.

Teachers were given a range of dates to choose from so that each teacher could pick their top choices of days to be videotaped. The dates of observations ranged from late September to early December in the fall and from early February to early May in the spring. Teachers' choices

were honored as often as possible, so most observations were conducted on days when teachers themselves chose to be observed. All teachers knew their observation dates ahead of time and gave verbal assent to be videotaped prior to the videographer beginning the recording. The goal was to observe the 90 minute reading block in its entirety. The average length of observations in the individual waves of data collection ranged from approximately 85 minutes to approximately 95 minutes. The standard deviations of the observation lengths ranged from 16 minutes to 26 minutes. Sometimes additional center and small group activities occurred outside of the normal 90 minute reading block and were not captured in the observation. Table 4 contains more details about the descriptive statistics for the observations.

Classroom Organization Rating Scale

In the current study, the rating scale was adapted from a holistic rating scale to a list of observable teacher or student behaviors that could be rated as individual items. Some examples include "The teacher uses behavior management consistently" and "Students obey the teacher immediately." Some of the original items such as "Students know what to do and what is expected of them" were difficult to describe using observable behaviors. The same item was also very similar to "The students seem to know what to do after the teacher has given directions." Other items, such as "Transitions are quick and efficient" turned out to be too subjective for consistent agreement between raters. Based on these observations, the list of items was narrowed down to include only the most readily observable and objective items. The final list of 24 items can be found in Appendix A.

Teacher behaviors were rated using a score of 0, 1, 2, and 3 for each of the 24 items. A score of 0 represented the category "No Evidence." Teachers received this score if the behavior was used successfully for less than 25% of the available or appropriate opportunities for its use.

For example, if there were six transitions in the tape and there only seemed to be a system for organizing students during one of the transitions (< 25%) the teacher would receive a score of "No Evidence" for the item "There is a system in place for facilitating students' transition from one part of the room to another." Similarly, a score of 1 represented the category of "Minimal Evidence" and was assigned if the behavior was observed for more than 25% and less than 50% of the available and appropriate opportunities, a score of 2 represented the category of "Some Evidence" and was assigned if the behavior was observed for more than 50% and less than 75% of the available and appropriate opportunities, and a score of 3 represented the category of "Sufficient Evidence" and was assigned if the behavior was observed for more than 75% of the available and appropriate opportunities. More information about the scoring criteria is available in Appendix A.

Rating Scale Procedures and Coder Training

All raters were certified teachers with experience teaching in the elementary grades. Three were graduate students in the college of education and two were retired teachers. One retired teacher was working as an adjunct in the teacher education department.

Raters watched the entire video for each teacher and kept notes during their viewing. They were trained to look for specific examples of evidence for each item. After viewing the video, the raters reviewed their notes and made final decisions about scoring. For each item, raters decided which score to assign based on the number of times that the teacher used the behavior out of the number of opportunities that the teacher had available to use the behavior. Each video was scored by at least two raters and scores were averaged for each item.

Raters were trained to score the classroom organization rating scale and reached adequate interrater reliability before beginning to score videos. First, raters participated in a half day

training focused on learning about the individual items on the scale. Then, they practiced scoring videos and comparing their answers. Disagreements were discussed until a consensus was reached and any ambiguities in the items were fleshed out in a decision-making document that raters kept as a reference while they were coding.

Percentage agreement and Kappa were monitored throughout the coding process. Each video was scored by at least two raters, so the percentage agreement and Kappa values reflect 100% of the observations that were scored, not just the ones that were used for training purposes. Kappa values ranging from 0.41 to 0.60 are considered to represent "moderate" agreement between raters and values ranging from 0.61 to 0.80 are considered to represent "substantial" agreement between raters (Landis & Koch, 1977). In this study the goal was for raters to achieve at least 85% agreement (kappa of approximately 0.65 to 0.70) for each observation with 75% agreement (kappa of approximately 0.55 to 0.60) being the minimum acceptable standard. Thus, raters were expected to achieve "moderate" to "substantial" levels of agreement as described by Landis and Koch (1977).

Table 5 shows the percent agreement and Kappa for each of the six waves of video observations. The average percent agreement for each wave of observations ranged from 80% agreement to 89% agreement. The overall average exact percent agreement was 85%. Likewise, the average Kappa values for each wave of observations ranged from 0.62 to 0.77, and the overall average Kappa value was 0.68. Thus, raters were successful at maintaining a "substantial" level of agreement on most observations. The number of observations falling below the minimum standard for reliability (75% agreement) is listed for each wave of observations in Table 5. The number of observations falling below the minimum standard for reliability ranged from 1 to 7 observations per wave with the higher numbers of ratings with less than 75%

agreement being in waves when new raters were being trained. These observations were scored by a third rater whenever possible.

Analytic Method

CFA for literacy and language measures. Confirmatory factor analysis was used to examine the appropriateness of a three factor model for the student assessments described above. The WIAT-III Word Reading subtest and the WJ-III Letter-Word Identification subtest observed scores were expected to load onto a word reading factor, the OWLS-II Listening Comprehension subtest and the WJ-III Oral Comprehension subtest observed scores were expected to load onto a listening comprehension factor, and the WIAT-III Reading Comprehension subtest and the WJ-III Passage Comprehension subtest observed scores were expected to load onto a reading comprehension factor. Standardized factor loadings and standard errors are reported for each set of observed scores.

Multiple fit indices are reported and interpreted below in order to assess the model's fit with the observed data. These include the chi-square test of model fit, Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). Fit indices for a single factor model, three two factor models, and a three factor model were examined and a chi-square difference test was used to select the best fitting model. Factor scores from the final model were extracted and used in the HLM analyses described below.

PCA for classroom organization rating scale. Principal components analysis (PCA) was used to examine how well the classroom organization rating scale items loaded onto one or more components for each wave of ratings. First, the scree plots were examined in order to determine the most likely number of meaningful components. The Oblimin with Kaiser

Normalization rotation method was used as the estimation method. Next, the component correlation matrix for each wave was examined for factors that were highly correlated. Then, the total variance explained by each component was reported as well as the loadings for each individual item.

Hierarchical linear modeling. Hierarchical Linear Modeling (HLM; multilevel modeling) in SPSS (IBM Corp., 2012) was used to address the research questions (Heck, Thomas, & Tabata, 2014; Raudenbush & Bryk, 2002). To address research question 1 "Is classroom organization in Grades 2 and 3 related to students' language and literacy skills?" cross-sectional HLM analyses were conducted separately for the second grade and third grade datasets. Classroom organization and its interactions with prior performance were included in the models as predictors of second and third grade students' word reading, reading comprehension, and listening comprehension. For each grade level, prior performance was estimated using the fall achievement scores and the spring scores were used as the outcome measures.

HLM models were built in several steps to address research question 2 "Does the classroom organization provided by first, second, and third grade teachers have a statistically significant relation to students' literacy and language achievement levels at the end of the third grade?" and research question 3 "Do the interactions of first, second, and third grade teachers' classroom organization with prior literacy and language performance predict students' achievement levels at the end of third grade?" These models were longitudinal in nature with six time points. This resulted in three-level models with time points nested within students and students nested within classrooms. At the highest level, the models were cross-classified because students were placed with at least three different teachers over the course of the study as they moved through the first, second, and third grades. Prior performance was estimated using the

achievement scores from the fall of first grade and scores from the spring of third grade were used as the outcome measures.

Restricted maximum likelihood estimation (REML) was used as the estimation method for the HLM models since REML uses only the variance components in the likelihood solution and does not include the point estimates of regression coefficients. Maximum likelihood estimation (ML) was not used because variance estimates produced are likely to be too small because the sample size for teachers in the current study is relatively small (Raudenbush & Bryk, 2002).

First, a set of intercept only models (unconditional models) was estimated. These models estimate the average achievement score for all occasions and all individuals, and the percentages of variance within individual, between individuals, and within classrooms are produced for each of the three outcome measures. The variance of the intercepts of student achievement at the end of the third grade was examined and reported as percentages of variance accounted for by each classroom placement. The second set of models included the classroom organization ratings for each teacher (level 3) as fixed effects. Finally, the interactions of classroom organization with prior literacy and language performance (cross-level interactions) in the first, second, and third grades were added to the models.

Power Analysis

For the models described above, power was simulated in R (R Development Core Team, 2008) using estimated effect sizes and interclass correlation coefficients (ICCs). The effect size for the classroom organization ratings was estimated to be 0.10 based on the correlation of classroom organization ratings with student outcomes in the first grade data. The first grade data were also used to estimate an ICC of 0.20 for the variation in literacy and language outcomes

within classrooms. A preliminary unconditional model with all six time points nested within students was used to estimate an ICC of 0.19 for the variation in literacy and language outcomes within student over all six time points. These effects and ICCs were entered into an R script (Z.W. Shen, personal communication, June 25, 2015) to simulate power for a three level model with time points nested within students and students nested within classrooms. The simulation was conducted for an average of 10 students per classroom and 30 classrooms per wave of data collection. These values are reflective of the data that were available for analysis in the current study since the average number of students per classroom ranged from 7 to 9 students and the number of classrooms per year ranged from 29 to 36. The above listed effect sizes and ICC's produce a simulated power of 0.71 to detect statistically significant classroom level effects for classroom organization.

CHAPTER FOUR

RESULTS

Descriptive Statistics and Preliminary Analyses

The student data and the classroom observation rating scale data were examined using the missing data analysis function in SPSS (IBM Corp., 2012). Of the original 373 student participants, 108 (28.95%) were missing data on at least one variable and about 12% of the total number of values in the dataset were missing. This was mainly due to attrition from the study. Little's missing completely at random (MCAR) test was used to determine whether there were any patterns in the data that were missing. The test was not significant ($\chi^2(414) = 456.33$, $p = .07$) indicating that the data were most likely missing completely at random. Data that were missing from the student measures dataset were imputed using the multiple imputation function in SPSS which generates multiple estimates of the missing values and then aggregates them into one estimate.

The percentage of missing data for the classroom organization rating scale was approximately 1% in year one, 18% in year two, and 25% in year three. The reason for the increase in the percentage of missing data as the study progressed is that teachers of retained students were not videotaped and thus these students did not have scores for the classroom organization rating scale. These students were deleted from the dataset.

Student Measures

Tables 5, 6, 7, and 8 display the descriptive statistics for students' raw scores and standard scores (based on the original and imputed data) on each of the measures for each of the six waves of data collection. Overall, the students' reading skills were in the average range compared to the norm sample (mean standard score of approximately 100). However there were

a few instances where the mean standard score was close to one standard deviation above the mean for the norm sample. For example, the WJ-III Oral Comprehension subtest average standard scores were 112.26, 113.74, and 114.21 in waves two, four, and six, respectively. The means after multiple imputation all differed from the original means by less than .60 raw score points and the standard deviations after multiple imputation all differed from the original standard deviations by less than .60 points. The skewness and kurtosis values (Table 9) indicated that overall the distributions of the variables were near enough to the normal distribution to be used without transformation. Raw scores were used in all subsequent analyses described below including confirmatory factor analyses using the student measures data and principal components analyses using the class organization rating scale data..

Correlations for the student measures' raw scores within each school year are presented in Tables 10, 11, and 12. The correlations of the WIAT-III Word Reading subtest with the WJ-III Letter-Word Identification subtest ranged from 0.87 to 0.90 within each wave with the strongest correlations in the first grade. The correlations of the OWLS-II Listening Comprehension subtest with the WJ-III Oral Comprehension subtest ranged from 0.48 to 0.62 within each wave with no discernible pattern across grade levels. The correlations of the WIAT-III Reading Comprehension subtest with the WJ-III Passage Comprehension subtest ranged from 0.60 to 0.78 within each wave. The correlations of the word reading measures with themselves across waves within each school year (i.e., WIAT-III Word Reading in wave 1 with WIAT-III Word Reading in wave 2) ranged from 0.84 to 0.89. For listening comprehension the range was from 0.62 to 0.64 and for reading comprehension the range was from 0.64 to 0.75.

Confirmatory factor analysis (CFA) was conducted using the original dataset in Mplus version 7.11 (Muthén & Muthén, 1998-2015) in order to examine the extent to which the models

of word reading, listening comprehension, and reading comprehension with two observed variables each exhibited measurement invariance and structural invariance across the six time points. Syntax for this analysis was adapted from an online example provided by Lesa Hoffman (2014). Differences in the chi-square values were used to make model comparisons. For each outcome, a configural invariance model was specified first with six correlated factors (i.e. the six time points) loading onto all six measures. For the purposes of model identification, the first indicator's loading was fixed to one and its intercept was fixed to zero for each factor. Factor variances, covariances, and means were estimated as well as the residual covariances of each indicator with itself across all six time points. The configural invariance model had mostly adequate fit to the data ($\chi^2(489,\ N = 295) = 1399.18$, p = .00, *CFI* = .92, *TLI* = .90, *RMSEA* = .08, *SRMR* = .06) according to the published standards (Hu & Bentler, 1999).

Next, successive models were estimated using parameter constraints in order to check for decreases in model fit due to measurement or structural invariance. The metric invariance model was first estimated to examine the equality of the unstandardized factor loadings across the six time points. Factor loadings for all six of the indicators (measures) of each factor (time point) were constrained to be equal across time and intercepts and residual variances were permitted to vary. The factor variance for the first time point was fixed to one and all others were allowed to vary. The metric invariance model fit significantly better than the configural model (χ^2 (25) = 41.46, p = .02) based on a chi-square difference test. The model modification indices indicated that the loadings of WJ-III Passage Comprehension for wave 1 and wave 4 were a source of misfit and should be freed. The partial metric invariance model with these two loadings freed fit significantly better than the full metric invariance model (χ^2 (2) = 22.33, p = .00). Overall, the

measures appeared to relate to the factors equally across time (i.e., the measures are invariant), with the WJ-III Passage Comprehension in waves one and four being the exceptions.

Confirmatory factor analysis (CFA) was also conducted using Mplus (Muthén & Muthén, 1998-2015) in order to reduce the number of observed variables to latent factors. This analysis included the imputed raw scores. One-factor, two-factor, and three-factor models were compared. The three-factor model included word reading, reading comprehension, and listening comprehension as three separate factors. All possible combinations of two-factor models were tested (i.e., word reading and reading comprehension as one factor and listening comprehension as a second factor; reading comprehension and listening comprehension as one factor and word reading as a separate factor; word reading and listening comprehension as one factor and reading comprehension as a separate factor). Chi-square difference tests showed that the three-factor model was superior to all other models in all six waves ($p < .001$).

After considering the fit indices for individual models and the results of the chi-square difference tests, the three-factor model was selected for data reduction purposes because it had the best model fit indices in general. Standardized loadings and residuals for the three-factor model can be found in Table 13. The three-factor model had adequate to excellent fit to the data in wave two ($\chi^2(6,\ N = 372) = 28.08$, $p = .00$, *CFI* = .99, *TLI* = .97, *RMSEA* = .10, *SRMR* = .02), wave three ($\chi^2(6,\ N = 372) = 43.19$, $p = .00$, *CFI* = .98, *TLI* = .95, *RMSEA* = .13, *SRMR* = .03), wave four ($\chi^2(6,\ N = 372) = 50.94$, $p = .00$, *CFI* = .98, *TLI* = .95, *RMSEA* = .13, *SRMR* = .03), wave five ($\chi^2(6,\ N = 372) = 62.63$, $p = .00$, *CFI* = .97, *TLI* = .91, *RMSEA* = .16, *SRMR* = .03), and wave six ($\chi^2(6,\ N = 372) = 43.45$, $p = .00$, *CFI* = .98, *TLI* = .94, *RMSEA* = .13, *SRMR* = .02; Hu & Bentler, 1999). The three-factor model had somewhat poorer fit to the data in wave one ($\chi^2(6,\ N = 372) = 237.31$, $p = .00$, *CFI* = .87, *TLI* = .68, *RMSEA* =

.32, *SRMR* = .11), but it still fit the data better than the one-factor model or any of the two-factor models. Factor scores were extracted from each three-factor model to be used in three separate HLM analyses.

Univariate analyses of the outcome measures and the bivariate analyses of the relations of organization with end of year performance revealed that there appeared to be four students who were outliers across all waves of data collection. These students had scores that were far below those of the other students for all three of the outcome measures. An example of the boxplots that were used as univariate checks for outliers can be found in Figure 1 which shows boxplots of word reading performance for each of the six time points. The outliers are located at the bottom of the plot area. An example of the scatterplots that were used as bivariate checks for outliers can be found in Figure 2 which depicts the relation between second grade classroom organization and second grade spring listening comprehension performance. The outliers are located at the bottom right corner of the plot and are indicated with a circle. In order to prevent these scores from biasing the results, these four students were removed from the data set prior to further analysis.

Classroom Organization Rating Scale

Table 14 shows the descriptive statistics and Cronbach's alpha values for the classroom organization rating scale. Interrater reliability for the classroom observation scale is reported in Table 4 and was discussed in detail in the method section. On average, teachers received approximately 61 out of the possible 72 points (average *SD* = 8.66) across all six waves of data collection. An average score of 61 indicates that the typical teacher was receiving about 85% of the total possible points. This means that the average teacher was rated as showing "some" or "sufficient" evidence for most of the items. This finding is consistent with other studies of

teacher observations in the United States that have found that the average of teacher quality measures tends to be in the medium to high range (e.g., Cadima et al., 2010; Curby et al., 2009; Kane & Staiger, 2012). The minimum score within wave ranged from 33.50 to 48.50 and the maximum score within wave ranged from 71.67 to 72.00. Cronbach's alpha within each wave ranged from 0.82 to 0.93.

Table 14 also displays the skewness and kurtosis values for the classroom organization ratings. The total raw scores for the classroom organization ratings appeared to be negatively skewed for all of the waves, except for waves three and six. The kurtosis values for the total raw scores of the classroom organization ratings in most of the waves showed that the ratings were fairly close to being normally distributed with the exceptions of waves two, four, and five which appeared to be more peaked than the normal distribution.

Table 15 shows the correlations of the classroom organization rating scale across waves. Within each year, the correlations of fall observations with spring observations were 0.65 (year 1), 0.44 (year 2), 0.70 (year 3; all significant at $p < .01$). Correlations between other combinations of waves were somewhat weaker with the significant ($p < .01$) values ranging from 0.19 to 0.35 and the non-significant values ranging from 0.04 to 0.12. Thus, as expected, the relations between ratings within the same school year were somewhat stronger than the relations of ratings from different school years, but the ratings from different school years were still moderately correlated in most cases. Since the classroom organization rating scale scores were significantly correlated within each year, the scores within each school year were averaged to create a single score for each year. HLM models were attempted without averaging and the models did not run because of multicollinearity between the observations conducted within the

same year. Therefore, although the correlation in year two was only 0.44, averaging was still used for consistency and model estimation purposes.

Principal components analysis (PCA) was used to explore the component structure of the classroom organization rating scale. Other methods, such as confirmatory factor analysis, may have resulted in factor scores with less error variance, but the sample of teachers in the current study was too small to provide adequate power for other types of factor analyses. Visual inspection of the principal component analysis scree plots for each wave of rating scale data showed that there were two possible components in each year of observation scores. The first component accounted for approximately 44 to 50% of the variance in each year and the second component accounted for approximately 12% of the variance in each year. The component matrix for each year of observations is presented in Table 16. For the first component, all of the items had loadings greater than 0.50 in at least one of the years except item 5 (the teacher elicits responses and ideas from students during the lesson), item 7 (the teacher incorporates students' examples and experiences into the lesson), and item 8 (children have many opportunities to talk appropriately during the lesson). These items were included in the PCA analysis, but contributed very little to the teachers' factor scores. For the second component, very few items within each year had loadings greater than 0.50 and the pattern of items with higher loadings across years did not make sense theoretically. Thus, the second component did not appear to be substantively meaningful and was not included in any further analyses. Factor scores were extracted for the first component and used as predictors in the multilevel models below.

The models used to address research question 1 were two-level designs with level one being the child level and level two being the classroom level. The models used to address research questions 2 and 3 were three-level repeated measures designs. Level one represented the

repeated measures that were nested within the child level (level two). Children were nested within classrooms (level three) and changed teachers each year, so level three was cross-classified.

Research Question 1: Is classroom organization in Grades 2 and 3 related to students' language and literacy skills?

Unconditional models were estimated separately for the second grade and third grade datasets. In second grade, classroom placement accounted for 16% of the total variance in word reading, 16% of the total variance in reading comprehension, and 8% of the total variance in listening comprehension. In third grade, classroom placement accounted for 9% of the total variance in word reading, 12% of the total variance in reading comprehension, and 12% of the total variance in listening comprehension. In comparison, Pilcher and Kim (2015) found that variation across classrooms accounted for 26% of the variance in spring word reading, 32% of the variance in students' spring reading comprehension, and 23% of the variance in students' spring listening comprehension.

Next, classroom organization was added to the models as a predictor of students' end of year achievement (Tables 17 and 18). There were no significant relations of classroom organization with students' end of year word reading, reading comprehension, or listening comprehension in second or third grade. In the first grade data (Pilcher & Kim, 2015) classroom organization was significantly and positively associated with students' reading comprehension and listening comprehension.

Finally, the interactions of prior performance (fall scores) with classroom organization ratings in the fall and spring were added to the models as predictors of end of year achievement in the second and third grades (Tables 19 and 20). No significant interactions of prior

performance with classroom organization were detected. In the first grade data (Pilcher & Kim, 2015), there was one significant interaction between prior performance and classroom organization that was predictive of students' spring word reading scores such that lower performing students had scores that were most similar to their peers when classroom organization was high.

Research Question 2: Does the classroom organization provided by first, second, and third grade teachers have a statistically significant relation to students' literacy and language achievement levels at the end of the third grade?

Unconditional models were estimated first in order to obtain the variance estimates for each level of the models. Variance in individuals' achievement estimates (level one) accounted for about 16% of the total variance in word reading, 18% of the total variance in reading comprehension, and 54% of the total variance in listening comprehension. The level two variance estimates, which are the variances in intercepts across individuals, represented 67% of the variance in word reading, 60% of the variance in reading comprehension, and 38% of the variance in listening comprehension. Wald Z tests were significant ($p < .001$) for the level one and level two variances in all three outcomes, which indicated that there was sufficient variation in intercepts across individuals and classrooms for further analyses (Heck, Thomas, & Tabata, 2014).

The classroom level variances accounted for much smaller percentages of the total variance estimates than the level one and level two variances. For the word reading outcome, the first grade classroom placement accounted for 0% of the total variance, the second grade classroom placement accounted for 14% of the total variance, and the third grade classroom placement accounted for 3% of the total variance. For the reading comprehension outcome the

estimates were 0%, 22%, and 0%, respectively, and for the listening comprehension outcome the estimates were 3%, 0%, and 5%, respectively, in Grades 1, 2, and 3.

Table 21 shows the bivariate correlations between the student measures factor scores and the classroom organization factor scores. Most correlations of student measures factor scores with classroom organization factor scores were non-significant ($p > .05$). There were significant correlations between the wave 1 listening comprehension factor and the year 2 classroom organization factor ($r = 0.13$, $p < .05$) as well as between the wave 6 listening comprehension factor and the year 3 classroom organization factor ($r = 0.14$, $p < .05$).

The second set of HLM models included the classroom organization factor scores and time variable as fixed effects (Table 22). Adding the quadratic function of time significantly improved model fit for listening comprehension ($p < .01$), but not for word reading or reading comprehension ($p > .05$). Therefore, the quadratic term was included in the model for listening comprehension, but not for word reading and reading comprehension. The effect of time was significant and positive for word reading and listening comprehension and the quadratic effect of time was significant and negative for listening comprehension. This indicates that children's growth in word reading and listening comprehension was generally positive, but listening comprehension growth slowed down over time. Although the effect of time was non-significant for reading comprehension, the descriptive statistics (Tables 5 and 6) show that students did show an appropriate amount of growth at each time point. The classroom organization factor scores for first grade, second grade, and third grade were not significant for any of the three outcome measures ($p > .05$).

Research Question 3: Do the interactions of first, second, and third grade teachers' classroom organization with prior literacy and language performance predict students' achievement levels at the end of third grade?

Next, the cross-level interactions of prior performance, defined as performance at wave one (fall of first grade), with classroom organization were added to the models. Prior performance was defined as performance at the beginning of first grade for all of the interactions rather than as performance at the beginning of each school year because this is the method used in related studies (Cadima et al., 2010; Curby et al., 2009) and because of Connor, Morrison, and Underwood's (2007) findings that the effects of instruction in first and second grade depended on the skills with which children began the first grade. Given the evidence in the literature on the persistence of students' growth trajectories based on their performance in the early grades (Jimerson, Egeland, & Teo, 1999; Juel, 1988; McClelland, Acock, & Morrison, 2006), it makes sense to use the simpler method of setting the first data point as the estimate for prior performance. The quadratic term for time was still only an improvement to the model for listening comprehension, and the effects of time and quadratic time remained unchanged.

Word reading. Table 23 displays the results of the model predicting word reading. The interaction of fall of first grade word reading performance and second grade classroom organization was a significant predictor of word reading at the end of third grade. Figure 3 shows that students whose word reading was below the mean in the fall of first grade had slightly higher, but nearly the same, word reading performance in the spring of the third grade when their second grade teachers provided lower levels of classroom organization as compared to when their second grade teachers provided higher levels of classroom organization. The opposite was true for students whose word reading scores were above the mean in the fall of the first grade.

They had much higher word reading scores in the spring of the third grade when their second grade teachers provided higher levels of classroom organization. A final model was estimated with only the significant interaction term included (Table 24).

Reading comprehension. Table 23 also displays the results of the model predicting reading comprehension. The interaction of reading comprehension performance in the fall of first grade and second grade classroom organization was a significant predictor of reading comprehension at the end of third grade. Figure 4 shows that students whose reading comprehension was below the mean in the fall of first grade had only slightly higher reading comprehension performance in the spring of the third grade when their second grade teachers provided higher levels of classroom organization as compared to when their second grade teachers provided lower levels of classroom organization. Students whose reading comprehension scores were above the mean in the fall of the first grade had much higher reading comprehension scores in the spring of the third grade when their second grade teachers provided higher levels of classroom organization. A final model was estimated with only the significant interaction term included (Table 24).

Listening comprehension. The results of the model predicting listening comprehension are shown in Table 23 as well. None of the interactions of listening comprehension prior performance and classroom organization were significant predictors of listening comprehension at the end of third grade. The final model for listening comprehension did not include any of the interaction effects (refer back to Table 22).

CHAPTER FIVE

DISCUSSION

The current study was designed to replicate and extend the findings of Pilcher and Kim (2015), which was a cross-sectional analysis of the relations of classroom organization ratings with student achievement in the first grade. The goals of the current study were to (1) replicate Pilcher and Kim's (2015) findings in second and third grade using the same dataset, (2) examine the longitudinal relations of classroom organization to each of three outcome measures, and (3) examine interactions of classroom organization to prior performance longitudinally as predictors of each of the three outcome measures. Classroom organization did not have a significant relation with any of the outcome measures, but two statistically significant interactions between classroom organization and prior performance were detected.

Research Question 1: Is classroom organization in Grades 2 and 3 related to students' language and literacy skills?

Pilcher and Kim (2015) found that classroom placement accounted for 23 to 32% of the variance in first graders' spring scores. The estimates for second and third grade are somewhat lower in the current study with second grade ranging from 9 to 18% of the variance and third grade ranging from 11 to 14% of the variance.

Second grade classroom organization was significantly and negatively related to students' end of year listening comprehension scores. This finding does not align with the findings from Pilcher and Kim (2015) nor does it align with previous related studies (Cadima et al., 2010; Connor et al., 2014; Foorman et al., 2006; Kane & Staiger, 2012).

The finding that nearly all of the relations of classroom organization with end of year achievement in second and third grade and all of the interactions of classroom organization with

prior performance were nonsignificant was not in agreement with previous findings. However, as discussed above, there appeared to be less variation in students' word reading, reading comprehension, and listening comprehension between classes in second and third grade than in first grade and this could explain why significant relations were not detected.

Research Question 2: Does the classroom organization provided by first, second, and third grade teachers have a statistically significant relation to students' literacy and language achievement levels at the end of the third grade?

The percentage of variance in students' literacy achievement accounted for at the classroom-level in most previous studies ranged from approximately 14% to 37% with most estimates in the lower end of the range (Carlisle, Kelcey, Berebitsky, & Phelps, 2011; Connor et al., 2014; Kane & Staiger, 2012). In a cross-sectional analysis of the first grade data from the current study, Pilcher and Kim (2015) found that the variance accounted for at the classroom-level ranged from 23 to 32%, depending on the outcome measure. Not surprisingly, the amount of variation accounted for by classroom level effects in the current study was small, ranging from 1% to 22% of the total variance. Third grade classroom-level effects accounted for the most variance in the listening comprehension model (5% of the total variance) and second grade classroom-level effects accounted for the most variance in the word reading and reading comprehension models (14% and 22%). First and third grade classroom-level effects accounted for the least amount of variance in the reading comprehension model (0%), second grade classroom-level effects accounted for the least amount of variance in the listening comprehension model (0%), and first grade classroom-level effects accounted for the least amount of variance in the word reading model (0%). The sum of the classroom-level effects

across all three grade levels was the greatest for reading comprehension (22% of the total variance) and the least for listening comprehension (8% of the total variance).

The bivariate correlations between the student measures factor scores and the classroom organization factor scores (Table 17) included two significant correlations, which indicated that there was some relation between student outcomes and classroom organization. Although classroom organization was expected to account for a relatively small amount of the total variance in students' literacy and language achievement, it was somewhat surprising that it did not have a significant main effect on any of the outcome measures for any of the three grade levels observed. The findings from the Connor et al. (2014) study, which used a similar rating scale as the current study with a sample of third grade teachers and students, included a marginally significant relation of teachers' rating scale scores with students' vocabulary and reading comprehension ($p = 0.06$). Pilcher and Kim (2015), which was conducted using the first grade data from the current study, found significant and positive relations of classroom organization ratings with reading comprehension and listening comprehension ($p < 0.05$). The Pilcher and Kim (2015) study was cross-sectional in nature, so there was no need for cross-classification to be included in the model. This could account for why significant main effects were found with the first grade sample but not when the entire longitudinal sample was included since cross-classification requires much more power to detect significant effects.

Other studies, using similar rating scales, found significant main effects for classroom organization and/or overall teacher effectiveness in a variety of areas (Kane & Staiger, 2012) including word reading, word attack skills, reading comprehension, spelling (Foorman et al., 2006) vocabulary, and print concepts (Cadima et al., 2010) in the first and second grades. Some studies have failed to find significant main effects of classroom organization on student

achievement in areas such as word reading, sound awareness, and math achievement in kindergarten and first grade (Cadima et al., 2010; Curby et al., 2009).

One likely explanation for the finding that classroom organization did not have a significant main effect in the current study is the small sample size of teachers at each grade level. The number of teachers included in the sample at each grade level ranged from 29 to 38 whereas previous studies had larger sample sizes. The MET study (Kane & Staiger, 2012) included almost 3,000 teachers, Cadima et al. (2010) included 64 teachers, and Foorman et al. (2006) included 53 teachers. The studies that did not find significant main effects (Curby et al., 2009; n = 36 kindergarten teachers and 37 first grade teachers) or found nearly significant main effects (Connor et al., 2014; n = 27 third grade teachers) had sample sizes that were comparable to the ones in the current study.

Research Question 3: Do the interactions of first, second, and third grade teachers' classroom organization with prior literacy and language performance predict students' achievement levels at the end of third grade?

Two significant interactions between classroom organization and prior performance were detected (Figures 3 and 4). Second grade teachers' classroom organization interacted with prior performance to predict students' third grade word reading and reading comprehension such that students whose scores were below the mean in the fall of the first grade had third grade scores that were very similar or only slightly higher when their second grade teachers provided higher levels of classroom organization. Students who began first grade with word reading and/or reading comprehension scores that were above the mean had third grade scores that were much higher when their teachers provided high levels of classroom organization in the second grade.

Although the direction of the interactions differs, these findings otherwise align with Connor, Morrison, and Underwood's (2007) finding that the instruction that students experienced in second grade was predictive of their word reading achievement while the instruction that the students experienced in first grade was not. The authors explained that the total amount of language arts instruction increased from first (63 minutes per day) to second grade (88 minutes per day) and that in second grade there was more of a focus on meaning-based instruction rather than the basic phonological awareness and letter-sound instruction that was observed in first grade.

Although the data are not included in the analyses above, the videos used in the current study were also coded for the amount of time spent on various types of reading and language arts instruction and non-instruction. An examination of the descriptive statistics from this alternative data source shows that the total time spent in reading and language arts instruction (with time spent in non-instruction removed) did increase from an average of about 65 minutes in the fall of first grade to an average of about 71 minutes in the fall of second grade. In addition, an increase in teachers' scores on the classroom organization rating scale was significantly and positively correlated with the amount of time spent on reading and language arts instruction in three out of the six waves of data collection (wave 1 $r = 0.44$, $p < 0.05$; wave 2 $r = 0.39$, $p < 0.05$; wave 3 $r = 0.25$, $p < 0.05$).

An examination of the areas of literacy development teachers chose to focus on during their language arts instruction shows that teachers in the first grade spent a greater proportion of their time on code-focused skills such as phonological awareness, grapheme-phoneme correspondence, decoding and encoding, and word reading fluency (36% in the fall and 30% in the spring) and a smaller proportion of their time on meaning-focused skills such as vocabulary

and comprehension (27% in the fall and 29% in the spring). Conversely, teachers in the second and third grades spent a smaller proportion of their time on code-focused skills (20%, 15%, 10%, and 10% in the fall and spring of second and third grades, respectively) and a greater proportion of their time on meaning-focused skills (41%, 43%, 55%, and 50% in the fall and spring of second and third grades, respectively). This descriptive analysis converges with Connor, Morrison, and Underwood's (2007) findings as well as providing a possible explanation for why, in the current study, there were differential effects for high versus low performing students in the second grade. That is, the focus of reading and language arts instruction changed in the second grade and children whose prior performance was low may not have been prepared for this change.

The interactions that were detected in the current study do not agree with the hypothesis that students with lower prior performance would have higher achievement if their teachers provided high levels of classroom organization (Cadima et al., 2010; Curby et al., 2009). The obvious explanation that higher levels of classroom organization were actually harmful to lower performing students does not seem to be logical nor does it align with previous studies. As mentioned above, increases in classroom organizational practices have been shown to be associated with increases in time spent on academic tasks, which is generally beneficial to students (Cameron et al., 2005; Connor, et al., 2004; Connor, et al., 2007; Connor et al., 2014; Foorman et al., 2006).

Previous analyses of the first grade data used in the current study (Pilcher & Kim, 2015) revealed significant interactions of classroom organization with prior performance that were in the expected direction. That is, increases in classroom organization were associated with higher performance on the word reading outcome for students in the lowest quartile of prior

performance. However, when the data were analyzed longitudinally these findings did not hold. Based on the findings of Pilcher and Kim (2015) and the descriptive analyses described above, it appears that high levels of classroom organization in the first grade are associated with increases in instructional time that is mainly focused on emergent literacy skills, which are associated with higher word reading achievement for students with low prior performance. However, improved word reading skills in first grade may not be sufficient to ensure success in the second grade if a minimum threshold of performance is not reached by the end of first grade.

Thus, a possible explanation for the findings of the current study is that students who do not achieve the necessary minimum threshold of performance in first grade are at such a disadvantage in comparison to their peers that they are unable to benefit from second grade level instruction, whether their teacher is highly organized or not. They are simply not ready for the more meaning-focused instruction that occurs in second grade. In contrast, the children who come to second grade with adequate skills are able to benefit from the increase in time spent on mostly meaning-focused instruction that occurs when their teacher is highly organized.

An alternative explanation could be that the amount of classroom organization that is provided may be more individualized rather than being a characteristic of the classroom as a whole. This problem has not been explored in research literature on classroom organization, mainly because an efficient method of collecting and scoring such data has yet to be developed. However, it is likely that students with lower prior performance receive more support from their teachers because of their greater need for increased time spent on academic instruction. So, the ratings of classroom organization in the current study may inaccurately estimate the amount of classroom organization that lower performing students received in comparison to their higher performing peers. Data on the classroom organization that individual students experienced rather

than the classroom as a whole may help to clarify the differential effects that were observed in the current study for higher performing versus lower performing students. The fact that classroom organization could not be coded at the level of the individual child is a major limitation in the current study as well as most, if not all, of the existing literature on this topic.

Limitations and Future Directions

As mentioned before, there are several limitations to the current study. Sample size is an issue for the classroom organization variable since taking cross-classification into account leaves each grade level with around 30 teacher participants. Another limitation that continues to be an issue in this body of literature is that there is currently no efficient way to score classroom organization at the individual child level. It should also be considered that these findings do not take into account other aspects of effective teacher-child interactions, such as emotional support and instructional support, which have been found to predict student achievement (Hamre et al., 2013).

More longitudinal studies are needed that make use of measures of classroom organization along with samples that include many more teachers from a variety of school settings. These studies should focus on whether the relations of aspects of effective teacher-child interactions (i.e., classroom organization, emotional support, instructional support) and their interactions with child characteristics change as children move into the upper elementary grades and beyond. There is also a need for more studies that examine the relations between classroom organization, time spent on academic tasks, and student achievement.

CHAPTER SIX

CONCLUSION

In this study, classroom organization was examined as a possible means by which to narrow the achievement gap between higher performing and lower performing students. During the primary grades students' growth trajectories are established, and, once established, the gap between lower performing students and higher performing students tends to widen over time (Jimerson, Egeland, & Teo, 1999; Juel, 1988; McClelland, Acock, & Morrison, 2006). The bioecological model of child development (Bronfenbrenner & Morris, 1998) states that characteristics of children's interactions with teachers and other school professionals, such as classroom organization, are some of the main factors that affect children's growth trajectories during their time in school. In the current study, classroom organization is hypothesized to be associated with higher student achievement because it facilitates an increase in the quantity (Cameron et al., 2005; Cameron et al., 2008; Foorman et al., 2006) and quality (Connor et al., 2014; Foorman et al., 2006) of time spent on instructional activities, which is associated with higher academic achievement.

Although the literature on classroom interactions contains an abundance of observational and correlational studies of high-quality teacher-child interactions, very few were longitudinal studies and nearly all of the existing studies focused on the preschool through first grade years. The current study represents an attempt to fill this gap by examining longitudinal data from the first through third grade years. In addition, interactions of students' performance at the beginning of the study with first, second, and third grade teachers' classroom organization were estimated as predictors of students' achievement at the end of third grade in order to examine the

possibility that high levels of classroom organization can help to close the achievement gap between high and low performing students.

In cross-sectional analyses of the second and third grade data classroom organization was not related to student achievement in the second grade or the third grade and no significant interactions of classroom organization with prior performance were detected. In the longitudinal analyses classroom organization was not significantly related to student achievement (word reading, reading comprehension, and listening comprehension) in any of the three grade levels, but the interaction of second grade classroom organization with first grade achievement levels was found to be a significant predictor of end of third grade word reading and reading comprehension. Students who began first grade with low performance on the measures of word reading and reading comprehension had similar or only slightly higher achievement at the end of third grade when their second grade teachers provided high levels of classroom organization. Students who began first grade with high performance on the measures of word reading and reading comprehension had much higher achievement at the end of third grade when their second grade teachers provided high levels of classroom organization.

The interactions that were found were not in the hypothesized direction. The most plausible explanation for these findings is that the focus of reading and language arts instruction is different beginning in the second grade, and children who are under-prepared for this shift in instructional focus are not able to benefit from increases in classroom organization in the same way that their higher achieving peers can benefit from it. Future directions were discussed including the design of more longitudinal studies with larger samples of teachers as well as finding ways to score aspects of classroom interactions at the individual child level.

Implications

The results of this study have important implications for practitioners. Mainly, being placed in a highly organized classroom is helpful to students who are on grade level or above grade level, but it does not necessarily appear to be beneficial to students who do not leave first grade with the necessary prerequisite skills for meaning-based reading instruction. That does not necessarily mean that classroom organization is not important. High levels of classroom organization seemed to benefit students with average and above average prior performance a great deal. It could be that classroom organization in combination with intensive remediation would be beneficial to vulnerable students who begin their schooling with fewer academic skills than their peers, but further work is needed in this area of research.

APPENDIX A

CLASSROOM ORGANIZATION RATING SCALE

Classroom Organization Rating Scale Items and Scoring Criteria	
Item 1	Procedures seem to have been established for managing student behavior.
Item 2	The teacher uses behavior management consistently.
Item 3	Students obey the teacher immediately.
Item 4	Instances of blatant misbehavior are rare.
Item 5	The teacher elicits responses and ideas from students during the lesson.
Item 6	The teacher frequently calls on a range of students.
Item 7	The teacher incorporates students' examples and experiences into the lesson.
Item 8	Students have many opportunities to talk appropriately during the lesson.
Item 9	The teacher redirects the students' behavior in respectful ways.
Item 10	The classroom can be described as a positive learning environment.
Item 11	The teacher has the lesson materials ready before the lesson begins.
Item 12	The students have the materials they need in order to complete the activities.
Item 13	There is a system in place for organizing students into groups during learning activities.
Item 14	There is a system in place for managing students' materials.
Item 15	Procedures are evident for daily classroom tasks.
Item 16	Students seem to be following a set of procedures for using classroom materials that have previously been taught.
Item 17	There is a system in place for facilitating students' transition from one part of the room to another.
Item 18	Transitions are quick and efficient.
Item 19	The students seem to know what to do after the teacher has given directions.
Item 20	The teacher does not have to repeat the directions.
Item 21	Students know what to do and what is expected of them.
Item 22	Classroom routine is evident.
Item 23	The pacing of the activities is appropriate for the students.
Item 24	The class time is used mainly for meaningful instruction.

Evidence of Indicators		
1	No Evidence	There is no indication that the behavior is present or the behavior is so rare that attempts to use it are ineffective. Available opportunities for using the behavior have not been utilized.
2	Minimal Evidence	Less than 50% of all appropriate opportunities to use the behavior are taken. An attempt has been made to incorporate the behavior, but it is usually ineffective.
3	Some Evidence	More than 50% of all appropriate opportunities to use the behavior are taken. The behavior is being used at a basic level. Most attempts to use the behavior are effective.
4	Sufficient Evidence	The behavior is obviously being utilized effectively on a regular basis. Almost all appropriate opportunities to use the behavior are taken.

APPENDIX B

HUMAN SUBJECTS APPROVAL MEMO

The Florida State University
Office of the Vice President For Research
Human Subjects Committee
Tallahassee, Florida 32306-2742
(850) 644-8673 · FAX (850) 644-4392

APPROVAL MEMORANDUM

Date: 12/7/2015

To: Heather Pilcher

Address:
Dept.: CURRICULM AND INSTRUCTION

From: Thomas L. Jacobson, Chair

Re: Use of Human Subjects in Research
Child by Classroom Organization Interactions in the First through Third Grades

The application that you submitted to this office in regard to the use of human subjects in the research proposal referenced above has been reviewed by the Human Subjects Committee at its meeting on 11/18/2015. Your project was approved by the Committee.

The Human Subjects Committee has not evaluated your proposal for scientific merit, except to weigh the risk to the human participants and the aspects of the proposal related to potential risk and benefit. This approval does not replace any departmental or other approvals, which may be required.

If you submitted a proposed consent form with your application, the approved stamped consent form is attached to this approval notice. Only the stamped version of the consent form may be used in recruiting research subjects.

If the project has not been completed by 11/16/2016 you must request a renewal of approval for continuation of the project. As a courtesy, a renewal notice will be sent to you prior to your expiration date; however, it is your responsibility as the Principal Investigator to timely request renewal of your approval from the Committee.

You are advised that any change in protocol for this project must be reviewed and approved by the Committee prior to implementation of the proposed change in the protocol. A protocol change/amendment form is required to be submitted for approval by the Committee. In addition,

federal regulations require that the Principal Investigator promptly report, in writing any unanticipated problems or adverse events involving risks to research subjects or others.

By copy of this memorandum, the Chair of your department and/or your major professor is reminded that he/she is responsible for being informed concerning research projects involving human subjects in the department, and should review protocols as often as needed to insure that the project is being conducted in compliance with our institution and with DHHS regulations.

This institution has an Assurance on file with the Office for Human Research Protection. The Assurance Number is FWA00000168/IRB number IRB00000446.

Cc: Young-Suk Kim, Advisor
HSC No. 2015.16182

APPENDIX C

TABLES

Table 1: Descriptive Information for Student Study Participants in Each Wave

Wave	Sample size	Mean age (SD)	% Male	% Caucasian	% African American	% Hispanic	% Asian	% Other race	% Receiving ESE services	% Receiving FRL	% ELL
1	351	6.4 (.53)	52.0 %	59.8 %	25.6 %	6.3 %	2.6 %	5.7 %	10.3 %	50.7 %	0.9 %
2	366	6.8 (.61)	51.9 %	59.6 %	26.0 %	6.0 %	2.5 %	6.0 %	10.4 %	52.1 %	0.8 %
3	327	7.3 (.52)	50.5 %	60.2 %	25.1 %	6.1 %	2.4 %	6.1 %	22.3 %	52.3 %	0.9 %
4	330	7.8 (.62)	50.9 %	60.0 %	25.2 %	6.4 %	2.4 %	6.4 %	22.4 %	51.5 %	0.9 %
5	311	8.3 (.54)	50.5 %	60.1 %	24.8 %	6.4 %	2.9 %	5.8 %	21.6 %	47.1 %	1.0 %
6	303	8.8 (.64)	50.5 %	60.1 %	24.4 %	6.6 %	3.0 %	5.9 %	18.8 %	46.9 %	1.0 %

Table 2: Descriptive Information about Teachers

Grade	Sample size	Mean years teaching (*SD*)	Mean years teaching current grade (*SD*)	Graduate Degree	State certification in a related area	Other Endorsements
1st	29	15.5 (13.1)	11.4 (10.7)	42.9 %	78.6 %	71.4 %
2nd	38	13.3 (11.0)	4.6 (6.2)	21.0 %	73.7 %	52.6 %
3rd	37	12.2 (7.7)	7.3 (6.0)	29.7 %	83.8 %	64.9 %

Table 3: Descriptive Statistics for the Video Observations

Wave	*n*	Mean Length (*SD*) (minutes)	Range of Dates Observed	Percentage Agreement	Kappa	*n* below 75% agreement	Coders
1	28	84.67 (24.44)	9/20 – 12/4	80%	0.62	6	1, 2, 3
2	29	91.32 (17.28)	2/5 – 5/13	81%	0.65	4	1, 2, 3
3	39	93.32 (16.45)	10/ 9 – 12/10	89%	0.77	1	1, 2, 3
4	38	92.09 (20.60)	2/11 – 4/18	85%	0.66	7	1, 2, 3, 4, 5
5	33	92.54 (17.45)	9/29 – 12/8	87%	0.70	2	2, 4, 5
6	34	95.14 (26.04)	2/13 – 5/6	87%	0.69	1	2, 4, 5

Table 4: Raw Scores and Standard Scores for Student Assessments

	Mean (*SD*)			Mean (*SD*)			Mean (*SD*)		
	Raw	SS	*a*	Raw	SS	*a*	Raw	SS	*a*
	Wave 1			Wave 2			Wave 3		
WIAT-III WR	14.14 (9.55)	100.15 (16.10)	0.95	23.02 (10.20)	104.80 (16.04)	0.95	29.14 (9.76)	104.94 (14.33)	0.95
WJ-III LWID	32.49 (6.75)	110.99 (13.54)	0.91	38.48 (6.61)	112.47 (12.89)	0.91	42.63 (6.05)	109.80 (11.25)	0.91
OWLS-II LC	69.82 (12.78)	100.64 (14.08)	0.93	78.40 (12.69)	105.39 (13.82)	0.93	85.14 (11.04)	108.14 (12.29)	0.93
WJ-III OC	13.03 (3.64)	103.79 (13.10)	0.74	16.18 (3.74)	112.26 (13.40)	0.75	16.58 (3.50)	107.06 (12.37)	0.74
WIAT-III RC	34.30 (14.34)	99.77 (14.61)	0.88	48.37 (13.03)	104.78 (14.50)	0.86	53.42 (11.55)	102.26 (13.18)	0.87
WJ-III PC	16.01 (4.50)	102.52 (14.09)	0.88	21.56 (3.90)	106.73 (12.45)	0.81	23.31 (4.48)	101.51 (11.27)	0.86
	Wave 4			Wave 5			Wave 6		
WIAT-III WR	34.36 (9.04)	106.40 (13.45)	0.95	39.11 (9.74)	105.85 (13.70)	0.95	42.94 (9.92)	106.19 (14.01)	0.95
WJ-III LWID	46.23 (6.03)	109.17 (11.31)	0.91	49.23 (6.53)	107.12 (11.48)	0.92	51.74 (5.76)	107.33 (10.61)	0.91
OWLS-II LC	90.44 (10.44)	110.40 (12.84)	0.92	94.20 (10.65)	110.37 (12.62)	0.93	98.21 (9.33)	111.79 (12.49)	0.91
WJ-III OC	19.32 (3.56)	113.74 (12.33)	0.75	19.38 (3.59)	107.90 (12.51)	0.79	21.99 (3.52)	114.21 (11.58)	0.76
WIAT-III RC	60.00 (10.49)	105.32 (12.58)	0.85	59.89 (10.92)	102.36 (13.42)	0.86	64.97 (10.68)	107.14 (14.28)	0.83
WJ-III PC	25.05 (3.20)	101.69 (9.43)	0.75	27.14 (4.20)	100.10 (11.11)	0.84	27.30 (3.53)	99.35 (10.06)	0.79

Note. WJ-III = Woodcock Johnson Tests of Achievement, Third Edition, LWID = Letter-Word Identification, PC = Passage Comprehension, OC = Oral Comprehension; WIAT-III = Wechsler Individual Achievement test, Third Edition, WR = Word Reading, RC = Reading Comprehension; OWLS-II = Oral and Written Language Scales, Second Edition, LC = Listening Comprehension.

Table 5: Raw Scores and Standard Scores for Student Assessments after Multiple Imputation

	Mean (*SD*)		Mean (*SD*)		Mean (*SD*)	
	Raw	SS	Raw	SS	Raw	SS
	Wave 1		Wave 2		Wave 3	
WIAT-III WR	14.01 (9.43)	99.91 (16.21)	23.03 (10.22)	104.76 (16.09)	27.60 (10.49)	102.70 (15.49)
WJ-III LWID	32.37 (6.67)	110.75 (13.71)	38.49 (6.62)	112.43 (12.91)	41.62 (6.56)	108.04 (12.13)
OWLS-II LC	69.79 (12.71)	100.44 (14.23)	78.35 (12.61)	105.36 (13.81)	83.41 (11.48)	106.40 (13.18)
WJ-III OC	13.04 (3.58)	103.76 (13.23)	16.15 (3.73)	112.21 (13.39)	16.18 (3.50)	105.61 (12.67)
WIAT-III RC	13.79 (8.52)	99.57 (14.76)	22.46 (7.98)	104.70 (14.56)	25.22 (8.64)	99.95 (14.20)
WJ-III PC	15.96 (4.46)	102.30 (14.36)	21.55 (3.90)	106.71 (12.46)	22.52 (4.77)	99.70 (11.94)
	Wave 4		Wave 5		Wave 6	
WIAT-III WR	33.02 (9.63)	104.76 (14.20)	37.01 (10.53)	103.37 (14.53)	40.76 (10.43)	103.80 (14.52)
WJ-III LWID	45.32 (6.44)	107.97 (11.69)	47.93 (6.83)	105.29 (11.82)	50.35 (6.32)	105.22 (11.38)
OWLS-II LC	89.15 (11.01)	109.00 (13.69)	92.60 (10.90)	108.72 (13.12)	96.44 (9.79)	111.79 (12.49)
WJ-III OC	19.03 (3.60)	112.80 (12.80)	18.79 (3.71)	106.07 (13.19)	21.53 (3.44)	112.96 (11.90)
WIAT-III RC	30.06 (6.92)	103.97 (13.28)	28.11 (8.13)	100.40 (14.15)	31.86 (6.83)	105.41 (14.75)
WJ-III PC	24.64 (3.30)	100.85 (9.53)	26.34 (4.37)	98.55 (11.41)	26.59 (3.59)	97.86 (10.36)

Note. WJ-III = Woodcock Johnson Tests of Achievement, Third Edition, LWID = Letter-Word Identification, PC = Passage Comprehension, OC = Oral Comprehension; WIAT-III = Wechsler Individual Achievement test, Third Edition, WR = Word Reading, RC = Reading Comprehension; OWLS-II = Oral and Written Language Scales, Second Edition, LC = Listening Comprehension.

Table 6: Skewness and Kurtosis for Student Assessments

	Wave 1		Wave 2		Wave 3		Wave 4		Wave 5		Wave 6	
	Skew (*SE*)	Kurtosis (*SE*)	Skew (*SE*)	Kurtosis (*SE*)	Skew (*SE*)	Kurtosis (*SE*)	Skew (*SE*)	Kurtosis (*SE*)	Skew (*SE*)	Kurtosis (*SE*)	Skew (*SE*)	Kurtosis (*SE*)
WIAT-III WR	1.06 (0.13)	0.33 (0.26)	0.40 (0.13)	-0.26 (0.26)	-0.05 (0.14)	-0.07 (0.27)	-0.10 (0.13)	0.51 (0.27)	-0.16 (0.14)	0.08 (0.28)	-0.32 (0.14)	-0.02 (0.28)
WJ-III LWID	0.65 (0.13)	0.40 (0.26)	0.37 (0.13)	0.16 (0.26)	0.21 (0.14)	0.44 (0.27)	-0.18 (0.13)	0.25 (0.27)	-0.18 (0.14)	1.01 (0.28)	-0.55 (0.14)	0.82 (0.28)
OWLS-II LC	-0.17 (0.13)	0.15 (0.26)	-0.27 (0.13)	-0.69 (0.26)	-0.35 (0.14)	0.40 (0.27)	-0.78 (0.13)	0.68 (0.27)	-0.35 (0.14)	0.07 (0.28)	-0.79 (0.14)	0.80 (0.28)
WJ-III OC	-0.18 (0.13)	-0.07 (0.26)	0.14 (0.13)	-0.23 (0.26)	-0.03 (0.14)	-0.12 (0.27)	-0.21 (0.13)	-0.38 (0.27)	-0.24 (0.14)	-0.01 (0.28)	-0.41 (0.14)	-0.45 (0.28)
WIAT-III RC	0.35 (0.13)	-0.86 (0.26)	-0.52 (0.13)	-0.40 (0.26)	-0.56 (0.14)	-0.40 (0.27)	-1.08 (0.13)	0.60 (0.27)	-0.94 (0.14)	0.72 (0.28)	-1.31 (0.14)	2.27 (0.28)
WJ-III PC	0.54 (0.13)	0.25 (0.26)	-0.33 (0.13)	0.40 (0.26)	-0.01 (0.14)	-0.62 (0.27)	-0.37 (0.13)	0.80 (0.27)	-0.38 (0.14)	0.31 (0.28)	-0.30 (0.14)	0.59 (0.28)

Note. WJ-III = Woodcock Johnson Tests of Achievement, Third Edition, LWID = Letter-Word Identification, PC = Passage Comprehension, OC = Oral Comprehension; WIAT-III = Wechsler Individual Achievement test, Third Edition, WR = Word Reading, RC = Reading Comprehension; OWLS-II = Oral and Written Language Scales, Second Edition, LC = Listening Comprehension.

Table 7: 1st Grade -- Correlations Between Student Assessments

		1	2	3	4	5	6	7	8	9	10	11
1	Wave 1 WIAT-III WR	1										
2	Wave 1 WJ-III LWID	.90**	1									
3	Wave 1 OWLS-II LC	.43**	.43**	1								
4	Wave 1 WJ-III OC	.38**	.37**	.58**	1							
5	Wave 1 WIAT-III RC	.75**	.76**	.59**	.54**	1						
6	Wave 1 WJ-III PC	.85**	.84**	.53**	.47**	.78**	1					
7	Wave 2 WIAT-III WR	.84**	.84**	.47**	.35**	.73**	.80**	1				
8	Wave 2 WJ-III LWID	.83**	.85**	.47**	.38**	.73**	.80**	.90**	1			
9	Wave 2 OWLS-II LC	.35**	.36**	.63**	.55**	.50**	.46**	.39**	.38**	1		
10	Wave 2 WJ-III OC	.38**	.35**	.60**	.64**	.57**	.45**	.39**	.38**	.57**	1	
11	Wave 2 WIAT-III RC	.63**	.65**	.63**	.50**	.76**	.67**	.72**	.70**	.55**	.57**	1
12	Wave 2 WJ-III PC	.67**	.69**	.54**	.42**	.70**	.72**	.79**	.77**	.49**	.49**	.77**

* $p < .01$ (2-tailed).

Note. WJ-III = Woodcock Johnson Tests of Achievement, Third Edition, LWID = Letter-Word Identification, PC = Passage Comprehension, OC = Oral Comprehension; WIAT-III = Wechsler Individual Achievement test, Third Edition, WR = Word Reading, RC = Reading Comprehension; OWLS-II = Oral and Written Language Scales, Second Edition, LC = Listening Comprehension.

Table 8: 2nd Grade -- Correlations Between Student Assessments

		1	2	3	4	5	6	7	8	9	10	11
1	Wave 3 WIAT-III WR	1										
2	Wave 3 WJ-III LWID	.89**	1									
3	Wave 3 OWLS-II LC	.37**	.42**	1								
4	Wave 3 WJ-III OC	.45**	.46**	.62**	1							
5	Wave 3 WIAT-III RC	.67**	.65**	.59**	.61**	1						
6	Wave 3 WJ-III PC	.76**	.76**	.51**	.57**	.74**	1					
7	Wave 4 WIAT-III WR	.89**	.86**	.36**	.42**	.58**	.72**	1				
8	Wave 4 WJ-III LWID	.85**	.84**	.38**	.44**	.60**	.72**	.89**	1			
9	Wave 4 OWLS-II LC	.35**	.37**	.62**	.57**	.58**	.54**	.36**	.35**	1		
10	Wave 4 WJ-III OC	.32**	.36**	.48**	.64**	.51**	.49**	.31**	.34**	.48**	1	
11	Wave 4 WIAT-III RC	.50**	.51**	.50**	.55**	.71**	.62**	.50**	.54**	.50**	.55**	1
12	Wave 4 WJ-III PC	.58**	.58**	.43**	.57**	.63**	.64**	.57**	.63**	.45**	.51**	.64**

* $p < .01$ (2-tailed).

Note. WJ-III = Woodcock Johnson Tests of Achievement, Third Edition, LWID = Letter-Word Identification, PC = Passage Comprehension, OC = Oral Comprehension; WIAT-III = Wechsler Individual Achievement test, Third Edition, WR = Word Reading, RC = Reading Comprehension; OWLS-II = Oral and Written Language Scales, Second Edition, LC = Listening Comprehension.

Table 9: 3rd Grade -- Correlations Between Student Assessments

		1	2	3	4	5	6	7	8	9	10	11
1	Wave 5 WIAT-III WR	1										
2	Wave 5 WJ-III LWID	.87**	1									
3	Wave 5 OWLS-II LC	.40**	.42**	1								
4	Wave 5 WJ-III OC	.42**	.43**	.59**	1							
5	Wave 5 WIAT-III RC	.59**	.61**	.55**	.54**	1						
6	Wave 5 WJ-III PC	.70**	.75**	.58**	.59**	.70**	1					
7	Wave 6 WIAT-III WR	.88**	.85**	.38**	.39**	.60**	.69**	1				
8	Wave 6 WJ-III LWID	.86**	.89**	.44**	.45**	.61**	.73**	.88**	1			
9	Wave 6 OWLS-II LC	.37**	.41**	.62**	.55**	.54**	.56**	.37**	.44**	1		
10	Wave 6 WJ-III OC	.34**	.38**	.55**	.64**	.51**	.55**	.34**	.40**	.57**	1	
11	Wave 6 WIAT-III RC	.57**	.59**	.53**	.55**	.75**	.67**	.57**	.60**	.53**	.51**	1
12	Wave 6 WJ-III PC	.62**	.68**	.51**	.49**	.63**	.73**	.69**	.68**	.49**	.57**	.60**

* $p < .01$ (2-tailed).

Note. WJ-III = Woodcock Johnson Tests of Achievement, Third Edition, LWID = Letter-Word Identification, PC = Passage Comprehension, OC = Oral Comprehension; WIAT-III = Wechsler Individual Achievement test, Third Edition, WR = Word Reading, RC = Reading Comprehension; OWLS-II = Oral and Written Language Scales, Second Edition, LC = Listening Comprehension.

Table 10: Literacy and Language Measures CFA -- Standardized Loadings and Residuals

	Wave 1		Wave 2		Wave 3		Wave 4		Wave 5		Wave 6	
	L (*SE*)	R (*SE*)	L (*SE*)	R (*SE*)	L (*SE*)	R (*SE*)	L (*SE*)	R (*SE*)	L (*SE*)	R (*SE*)	L (*SE*)	R (*SE*)
Word reading												
WIAT-III WR	.98 (.01)	.04 (.02)	.96 (.01)	.07 (.02)	.95 (.01)	.10 (.02)	.92 (.01)	.16 (.03)	.93 (.01)	.14 (.03)	.93 (.01)	.14 (.02)
WJ-III LWID	.89 (.01)	.20 (.02)	.90 (.01)	.19 (.02)	.92 (.01)	.16 (.02)	.96 (.01)	.08 (.02)	.90 (.02)	.20 (.03)	.93 (.01)	.13 (.02)
Reading comprehension												
WIAT-III RC	.88 (.01)	.23 (.03)	.86 (.02)	.27 (.03)	.87 (.02)	.24 (.03)	.86 (.02)	.27 (.03)	.84 (.02)	.30 (.03)	.78 (.02)	.40 (.04)
WJ-III PC	.91 (.01)	.17 (.02)	.90 (.02)	.20 (.03)	.91 (.01)	.17 (.02)	.85 (.02)	.27 (.03)	.90 (.01)	.18 (.03)	.85 (.02)	.28 (.03)
Listening comprehension												
OWLS-II LC	-.64 (.26)	.59 (.33)	.74 (.04)	.45 (.05)	.78 (.03)	.39 (.05)	.70 (.04)	.51 (.05)	.77 (.03)	.41 (.05)	.69 (.04)	.52 (.05)
WJ-III OC	.09 (.06)	.99 (.01)	.76 (.04)	.43 (.05)	.82 (.03)	.33 (.05)	.79 (.03)	.38 (.05)	.77 (.03)	.40 (.05)	.82 (.03)	.33 (.05)

Note. WJ-III = Woodcock Johnson Tests of Achievement, Third Edition, LWID = Letter-Word Identification, PC = Passage Comprehension, OC = Oral Comprehension; WIAT-III = Wechsler Individual Achievement test, Third Edition, WR = Word Reading, RC = Reading Comprehension; OWLS-II = Oral and Written Language Scales, Second Edition, LC = Listening Comprehension; L = standardized loading; R = residual

Table 11: Descriptive Statistics for the Classroom Organization Rating Scale Raw Scores

Wave	*n*	Mean (*SD*)	Min.	Max.	*a*	Skew (*SE*)	Kurtosis (*SE*)
1	28	60.21 (9.58)	34.00	72.00	0.91	-0.92 (.45)	0.34 (0.87)
2	29	60.70 (9.65)	35.50	71.67	0.85	-0.97 (0.43)	0.98 (0.85)
3	36	61.87 (6.32)	48.50	72.00	0.82	-0.04 (0.39)	-0.89 (0.77)
4	37	62.94 (9.57)	33.50	72.00	0.93	-1.44 (0.39)	1.40 (0.76)
5	36	62.83 (8.44)	37.83	72.00	0.92	-1.52 (0.39)	2.09 (0.77)
6	37	61.73 (8.40)	42.00	72.00	0.91	-0.86 (0.39)	-0.23 (0.76)

Table 12: Classroom Organization Rating Scale Correlations

		1	2	3	4	5
1	Wave 1	1				
2	Wave 2	.65**	1			
3	Wave 3	.35**	.32**	1		
4	Wave 4	.23**	.26**	.44**	1	
5	Wave 5	.24**	.19**	.04	.12	1
6	Wave 6	.32**	.30**	.11	.20**	.70**

** $p < .01$ (2-tailed).

Table 13: Component Matrix for PCA with Classroom Organization Rating Scale Items

Year	1		2		3	
Components	1	2	1	2	1	2
Item 1	0.17	0.00	0.63	-0.21	0.56	-0.49
Item 2	0.83	-0.13	0.82	-0.16	0.81	-0.33
Item 3	0.86	-0.31	0.85	-0.05	0.76	-0.40
Item 4	0.87	-0.17	0.86	-0.18	0.83	-0.37
Item 5	0.10	0.72	0.25	0.74	0.45	0.54
Item 6	0.42	0.35	0.52	0.70	0.51	0.23
Item 7	-0.05	0.79	0.27	0.65	0.28	0.41
Item 8	0.29	0.84	0.40	0.40	0.42	0.65
Item 9	0.80	0.38	0.67	0.06	0.80	0.03
Item 10	0.84	0.27	0.83	-0.07	0.65	0.30
Item 11	0.44	0.24	0.43	-0.54	0.75	0.05
Item 12	0.80	0.05	0.17	-0.24	0.41	0.01
Item 13	0.41	-0.29	0.40	-0.22	0.52	0.53
Item 14	0.65	0.23	0.66	-0.35	0.52	0.39
Item 15	0.82	-0.05	0.81	-0.08	0.82	-0.10
Item 16	0.84	0.08	0.77	-0.17	0.56	-0.11
Item 17	0.76	-0.13	0.60	-0.35	0.71	-0.40
Item 18	0.71	-0.27	0.73	-0.38	0.73	-0.49
Item 19	0.85	-0.13	0.81	0.25	0.85	0.02
Item 20	0.84	-0.16	0.86	0.31	0.88	-0.13
Item 21	0.78	-0.37	0.83	0.10	0.83	-0.18
Item 22	0.87	-0.06	0.72	0.11	0.88	0.10
Item 23	0.92	0.16	0.74	0.04	0.75	0.45
Item 24	0.82	0.09	0.54	0.37	0.82	0.34

Table 14: Second and Third Grade Cross-Sectional Hierarchical Linear Modeling Results

	Word Reading			Reading Comprehension			Listening Comprehension		
	Coeff.	*SE*	*p*	Coeff.	*SE*	*p*	Coeff.	*SE*	*p*
Second Grade without Interaction Effects									
Intercept	0.43	0.24	0.08	0.40	0.21	0.07	0.18	0.10	0.09
Fall scores	0.09	0.23	0.70	-0.32	0.21	0.14	-0.22	0.11	0.05
Classroom organization	0.89	0.03	<0.001	0.64	0.03	<0.001	0.74	0.04	<0.001
Second Grade with Interaction Effects									
Intercept	0.42	0.24	0.09	0.40	0.21	0.07	0.18	0.10	0.10
Fall Scores	-0.03	0.24	0.90	-0.33	0.22	0.15	-0.22	0.11	0.05
Classroom organization	0.89	0.03	<0.001	0.64	0.03	<0.001	0.74	0.04	<0.001
Classroom Organization * Prior performance	0.05	0.03	0.11	0.00	0.03	0.90	0.00	0.04	0.98
Third Grade without Interaction Effects									
Intercept	0.14	0.17	0.43	0.04	0.15	0.77	-0.01	0.09	0.93
Fall scores	-0.18	0.14	0.22	0.01	0.14	0.94	0.17	0.09	0.07
Classroom organization	0.86	0.03	<0.001	0.73	0.02	<0.001	0.82	0.04	<0.001
Third Grade with Interaction Effects									
Intercept	0.14	0.17	0.43	0.06	0.16	0.71	0.00	0.09	0.98
Fall Scores	-0.16	0.17	0.34	0.08	0.15	0.60	0.18	0.09	0.06
Classroom organization	0.86	0.03	<0.001	0.73	0.02	<0.001	0.82	0.04	<0.001
Classroom Organization * Prior performance	-0.01	0.03	0.83	-0.03	0.02	0.23	-0.02	0.04	0.55

Table 15: Factor Score Correlations

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. W1 WI	1																			
2. W1 RC	0.95**	1																		
3. W1 LC	-0.67**	-0.83**	1																	
4. W2 WI	0.87**	0.86**	-0.65**	1																
5. W2 RC	0.80**	0.86**	-0.75**	0.92**	1															
6. W2 LC	0.50**	0.66**	-0.71**	0.55**	0.82**	1														
7. W3 WI	0.81**	0.81**	-0.62**	0.91**	0.86**	0.55**	1													
8. W3 RC	0.74**	0.82**	-0.72**	0.83**	0.90**	0.74**	0.92**	1												
9. W3 LC	0.49**	0.64**	-0.69**	0.53**	0.72**	0.82**	0.60**	0.85**	1											
10. W4 WI	0.75**	0.77**	-0.61**	0.87**	0.83**	0.54**	0.91**	0.84**	0.54**	1										
11.W4 RC	0.66**	0.76**	-0.71**	0.74**	0.85**	0.77**	0.78**	0.87**	0.77**	0.84**	1									
12. W4 LC	0.46**	0.61**	-0.66**	0.49**	0.69**	0.80**	0.53**	0.73**	0.81**	0.55**	0.90**	1								
13. W5 WI	0.74**	0.76**	-0.60**	0.86**	0.84**	0.58**	0.90**	0.86**	0.60**	0.92**	0.81**	0.57**	1							
14. W5 RC	0.68**	0.77**	-0.68**	0.78**	0.87**	0.76**	0.84**	0.91**	0.78**	0.83**	0.89**	0.75**	0.92**	1						
15. W5 LC	0.53**	0.66**	-0.65**	0.59**	0.77**	0.82**	0.65**	0.82**	0.85**	0.63**	0.83**	0.82**	0.70**	0.93**	1					
16. W6 WI	0.69**	0.71**	-0.57**	0.81**	0.81**	0.56**	0.87**	0.83**	0.58**	0.92**	0.81**	0.56**	0.94**	0.87**	0.67**	1				
17. W6 RC	0.65**	0.73**	-0.64**	0.75**	0.83**	0.72**	0.83**	0.88**	0.75**	0.85**	0.88**	0.72**	0.88**	0.92**	0.82**	0.93**	1			
18. W6 LC	0.44**	0.58**	-0.59**	0.49**	0.67**	0.78**	0.57**	0.74**	0.81**	0.55**	0.77**	0.79**	0.60**	0.78**	0.83**	0.62**	0.86**	1		
19. Y1 ORG	0.01	0.02	-0.04	0.01	-0.01	-0.01	0.07	0.06	0.05	0.04	0.05	0.04	0.03	0.02	0.02	0.08	0.09	0.08	1	
20. Y2 ORG	-0.05	-0.07	0.13*	-0.07	-0.10	-0.11	-0.01	-0.01	0.01	-0.03	-0.10	-0.11	-0.06	-0.09	-0.10	-0.01	-0.01	-0.01	0.35**	1
21. Y3 ORG	0.02	0.02	0.05	0.02	0.02	0.01	0.10	0.10	0.07	-0.02	0.01	0.03	0.04	0.07	0.09	0	0.07	0.14*	0.36**	0.25**

** $p < .01$ (2-tailed).
* $p < .05$ (2-tailed).

Note. W1 – W6 = Wave 1 through Wave 6, WR = Word Reading, RC = Reading Comprehension, LC = Listening Comprehension, ORG = Classroom Organization

Table 16: Longitudinal Hierarchical Linear Modeling Results for Word Reading, Reading Comprehension, and Listening Comprehension

	Word Reading			Reading Comprehension			Listening Comprehension		
	Coeff.	*SE*	*p*	Coeff.	*SE*	*p*	Coeff.	*SE*	*p*
Intercept	1.25	0.52	0.02	1.73	0.57	0.01	-0.67	0.23	0.00
Time	0.08	0.03	0.02	-0.03	0.04	0.43	0.62	0.11	0.00
Quadratic time							-0.07	0.02	0.00
Y1 classroom organization	0.42	0.42	0.34	0.41	0.40	0.31	0.09	0.14	0.53
Y2 classroom organization	0.01	0.52	0.99	-0.09	0.60	0.89	-0.06	0.12	0.63
Y3 classroom organization	0.02	0.42	0.97	0.00	0.40	0.99	0.18	0.15	0.23

Note. Y1 = Year 1, Y2 = Year 2, Y3 = Year 3.

Table 17: Longitudinal Hierarchical Linear Modeling Results with Interaction Effects for Word Reading, Reading Comprehension, and Listening Comprehension

	Word Reading			Reading Comprehension			Listening Comprehension		
	Coeff.	*SE*	*p*	Coeff.	*SE*	*p*	Coeff.	*SE*	*p*
Intercept	1.15	0.49	0.03	1.69	0.56	0.01	-0.67	0.23	0.00
Time	0.08	0.03	0.02	-0.03	0.04	0.43	0.62	0.11	0.00
Quadratic time							-0.07	0.02	0.00
Y1 classroom organization	0.46	0.39	0.25	0.44	0.39	0.27	0.10	0.14	0.50
Y2 classroom organization	0.03	0.51	0.95	-0.18	0.59	0.76	-0.10	0.12	0.43
Y3 classroom organization	-0.08	0.40	0.84	-0.04	0.40	0.91	0.17	0.15	0.27
Y1org * Prior performance	0.10	0.07	0.13	0.05	0.06	0.40	-0.14	0.11	0.21
Y2org * Prior performance	0.14	0.07	0.03	0.17	0.06	0.01	-0.09	0.10	0.35
Y3org * Prior performance	0.10	0.05	0.07	0.06	0.05	0.26	0.01	0.10	0.95

Note. Y1 = Year 1, Y2 = Year 2, Y3 = Year 3.

Table 18: Final Longitudinal Hierarchical Linear Modeling Results for Word Reading, Reading Comprehension, and Listening Comprehension

	Word Reading			Reading Comprehension		
	Coeff.	*SE*	*p*	Coeff.	*SE*	*p*
Intercept	1.28	0.53	0.02	1.77	0.58	0.01
Time	0.08	0.03	0.02	-0.03	0.04	0.43
Quadratic time						
Y1 classroom organization	0.43	0.41	0.32	0.40	0.39	0.31
Y2 classroom organization	-0.08	0.54	0.89	-0.24	0.61	0.70
Y3 classroom organization	0.10	0.41	0.80	0.07	0.39	0.86
Y1org * Prior performance						
Y2org * Prior performance	0.20	0.06	0.00	0.20	0.06	0.00
Y3org * Prior performance						

Note. Y1 = Year 1, Y2 = Year 2, Y3 = Year 3.

APPENDIX D

FIGURES

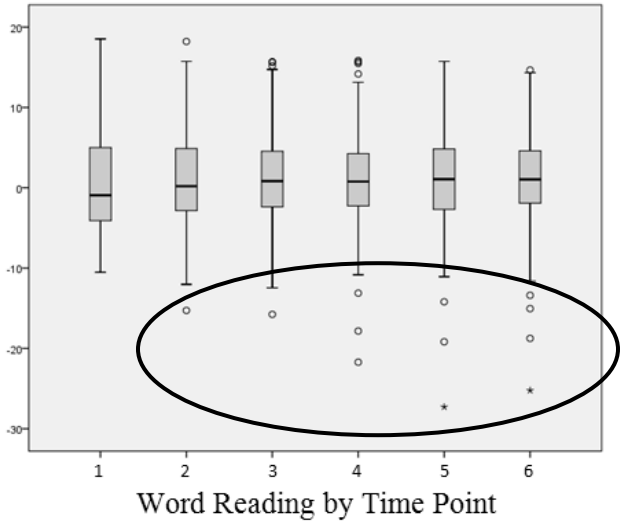

Figure 1: Example of Boxplot Graph Used for Univariate Outlier Analysis

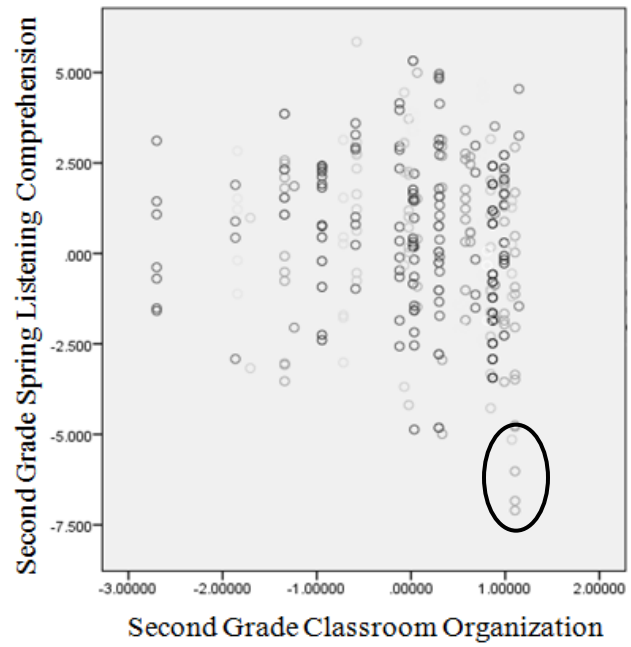

Figure 2: Example of Scatterplot Used for Bivariate Outlier Analysis

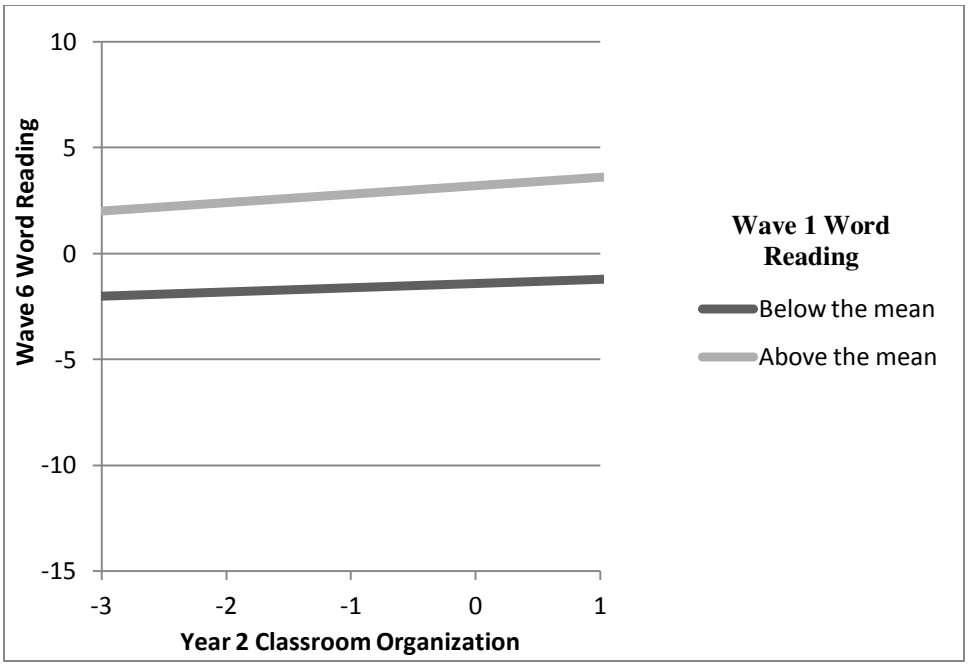

Figure 3: Interaction of Year 2 Classroom Organization with Word Reading Prior Performance

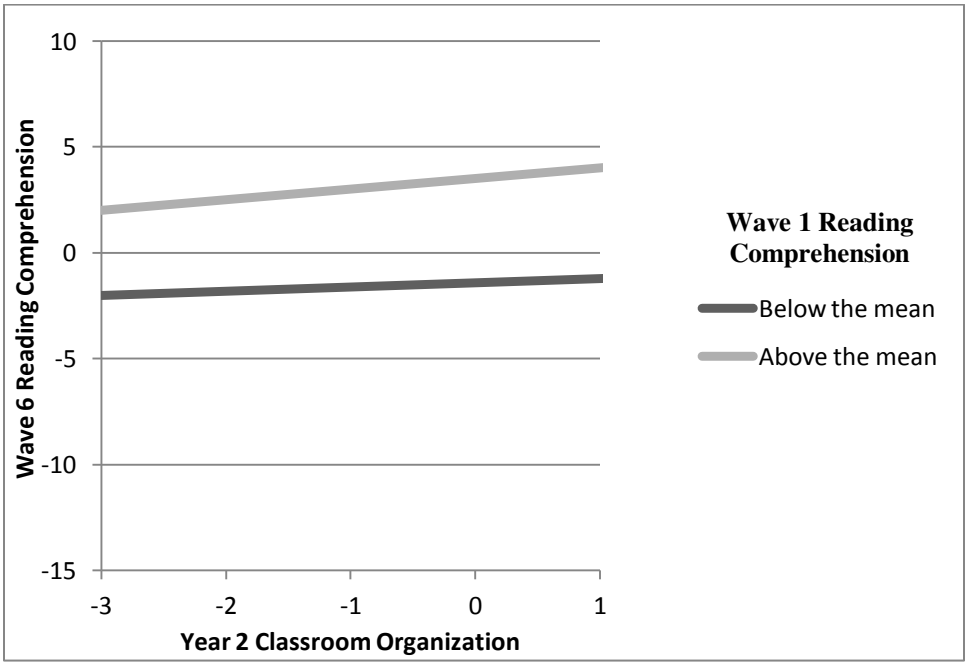

Figure 4: Interaction of Year 2 Classroom Organization with Reading Comprehension Prior Performance

REFERENCES

Anderson, L. M., Evertson, C. M., & Brophy, J. E. (1979). An experimental study of effective teaching in first-grade reading groups. *The Elementary School Journal, 79*(4), 193-223. doi:10.1086/461151

Bohn, C. M., Roehrig, A. D., & Pressley, M. (2004). The first days of school in the classrooms of two more effective and four less effective primary-grades teachers. *The Elementary School Journal*, *104*(4), 269-287. doi:10.1086/499753

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In W. Damon & R.M. Lerner (Eds.), *Handbook of child psychology:Vo 1: Theoretical models of human development* (pp. 993-1028). New York: Wiley.

Brown, J. L., Jones, S. M., LaRusso, M. D., & Aber, J. L. (2010). Improving classroom quality: Teacher influences and experimental impacts of the 4Rs program. *Journal of Educational Psychology*, *102*(1), 153-167. doi:10.1037/a0018160

Cadima, J., Leal, T., & Burchinal, M. (2010). The quality of teacher–student interactions: Associations with first graders' academic and behavioral outcomes. *Journal of School Psychology*, *48*(6), 457-482. doi:10.1016/j.jsp.2010.09.001

Cameron, C. E., Connor, C. M., & Morrison, F. J. (2005). Effects of variation in teacher organization on classroom functioning. *Journal of School Psychology, 43*(1), 61-85. doi:10.1016/j.jsp.2004.12.002

Cameron, C. E., Connor, C. M., Morrison, F. J., & Jewkes, A. M. (2008). Effects of classroom organization on letter–word reading in first grade. *Journal of School Psychology, 46*(2), 173-192. doi:10.1016/j.jsp.2007.03.002

Carlisle, J., Kelcey, B., Berebitsky, D., & Phelps, G. (2011). Embracing the complexity of instruction: a study of the effects of teachers' instruction on students' reading comprehension. *Scientific Studies of Reading*, *15*(5), 409-439. doi: 10.1080/10888438.2010.497521

Carrow-Woolfolk, E. (2011). Oral and written language scales (OWLS). Los Angeles, CA: Western Psychological Services.

Catts, H., Adlof, S., & Ellis Weismer, S. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research, 49*, 278-293.

Chatterji, M. (2006). Reading achievement gaps, correlates, and moderators of early reading achievement: Evidence from the early childhood longitudinal study (ECLS) kindergarten to first grade sample. *Journal of Educational Psychology*, *98*(3), 489-507. doi:10.1037/0022-0663.98.3.489

Connor, C. M., Morrison, F. J., & Katch, L. E. (2004). Beyond the reading wars: Exploring the effect of child-instruction interactions on growth in early reading. *Scientific Studies of Reading*, *8*(4), 305-336. doi:10.1207/s1532799xssr0804_1

Connor, C., Morrison, F., & Underwood, P. (2007). A second chance in second grade: The independent and cumulative impact of first- and second-grade reading instruction and students' letter-word reading skill growth. *Scientific Studies of Reading*, *11*(3), 199-233. doi:10.1080/10888430701344314

Connor, C. M., Spencer, M., Day, S. L., Giuliani, S., Ingebrand, S. W., McLean, L., & Morrison, F. J. (2014). Capturing the complexity: Content, type, and amount of instruction and quality of the classroom learning environment synergistically predict third graders' vocabulary and reading comprehension outcomes. *Journal of Educational Psychology*, *106*(3), 762-778. doi:10.1037/a0035921

Curby, T., Brock, L., & Hamre, B. (2013). Teachers' emotional support consistency predicts children's achievement gains and social skills. *Early Education and Development*, *24*(3), 292-309. doi:10.1080/10409289.2012.665760

Curby, T. W., Rimm-Kaufman, S. E., & Ponitz, C. C. (2009). Teacher-child interactions and children's achievement trajectories across kindergarten and first grade. *Journal of Educational Psychology*, *101*(4), 912-925. doi:10.1037/a0016647

Danielson, C. (2013). The Framework for Teaching Evaluation Instrument. The Danielson Group. Retrieved from http://www.danielsongroup.org/download/?download=448

Downer, J., Hamre, B., & Sabol, T. (2010). Teacher-child interactions in the classroom: Toward a theory of within- and cross-domain links to children's developmental outcomes. *Early Education & Development*, *21*(5), 699-723. doi:10.1080/10409289.2010.497453

Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 392-431). New York : London: Macmillan.

Foorman, B. R., Schatschneider, C., Eakin, M. N., Fletcher, J. M., Moats, L. C., & Francis, D. J. (2006). The impact of instructional practices in grades 1 and 2 on reading and spelling achievement in high poverty schools. *Contemporary Educational Psychology, 31*(1), 1-29. doi:10.1016/j.cedpsych.2004.11.003

Guo, Y., Sun, S., Breit-Smith, A., Morrison, F. J., & Connor, C. M. (2014). Behavioral engagement and reading achievement in elementary-school-age children: A longitudinal cross-lagged analysis. *Journal of Educational Psychology*. *107*(2), 332. http://dx.doi.org/10.1037/a0037638

Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher–child interactions: Associations with preschool children's development. *Child Development*, *85*(3), 1257-1274. doi:10.1111/cdev.12184

Hamre, B. K., Rivers, S. E., Brackett, M. A., Hamagami, A., Pianta, R. C., Downer, J. T., . . . Atkins, M. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *Elementary School Journal, 113*(4), 461-487. doi:10.1086/669616

Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, *72*(2), 625-638. doi:10.1111/1467-8624.00301

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development, 76*(5), 949-949. doi:10.1111/j.1467-8624.2005.00889.x

Hamre, B. K., & Pianta, R. C. (2007). Learning opportunities in preschool and early elementary classrooms. In R. C. Pianta, M. J. Cox, & K. Snow (Eds.), *School readiness and the transition to school* (pp. 49–84). Baltimore, MD: Brookes.

Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., LoCasale-Crouch, J., Downer, J. T., . . . Scott-Little, C. (2012). A course on effective teacher-child interactions: Effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal, 49*(1), 88-123. doi:http://dx.doi.org.proxy.lib.fsu.edu/10.3102/0002831211434596

Heck, R. H., Thomas, S., & Tabata, L. N. (2014). *Multilevel and longitudinal modeling with IBM SPSS*. New York; London: Routledge.

Hoffman, L. (2014). *Longitudinal measurement invariance in CFA using Mplus.* Retrieved from http://www.lesahoffman.com/948/index.html

Hoover, W. & Gough, P. (1990). The simple view of reading. *Reading & Writing*, *2*, 127-160.

Houghton Mifflin Harcourt. (2008). *Storytown.* San Diego, CA: Author.

Houghton Mifflin Harcourt. (2014). *Journey's Common Core,* San Diego, CA: Author.

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

IBM SPSS Statistics for Windows, (Version 21.0) [Computer software]. Armonk, NY: IBM Corp.

Jimerson, S., Egeland, B., & Teo, A. (1999). A longitudinal study of achievement trajectories: Factors associated with change. *Journal of Educational Psychology*, *91*(1), 116-126. doi:10.1037/0022-0663.91.1.116

Joshi, R. M., Tao, S., Aaron, P. G., & Quiroz, B. (2012). Cognitive component of componential model of reading applied to different orthographies. *Journal of Learning Disabilities, 45*, 480-486.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, *80*(4), 437. doi:10.1037/0022-0663.80.4.437

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). Seattle, WA: Bill & Melinda Gates Foundation.

Kim, Y.-S. (2015). Language and cognitive predictors of text comprehension: Evidence from multivariate analysis. *Child Development*, *86*, 128-144.

Kim, Y.-S. G., & Petscher, Y. (in press). Prosodic sensitivity and reading: An investigation of pathways of relations using a latent variable approach. *Journal of Educational Psychology.*

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174.

Li Grining, C., Raver, C. C., Champion, K., Sardin, L., Metzger, M., & Jones, S. M. (2010). Understanding and improving classroom emotional climate and behavior management in the "real world": The role of head start teachers' psychosocial stressors. *Early Education & Development, 21*(1), 65-94. doi:10.1080/10409280902783509

Marzano, R. J. (2007). *The art and science of teaching: A comprehensive framework for effective instruction*. Alexandria, VA: The Association for Supervision and Curriculum Development (ASCD).

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

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D.. . Howes, C. (2008). Measures of classroom quality in prekindergarten and childrens development of academic, language, and social skills. *Child Development*, *79*(3), 732-749. doi:10.1111/j.1467-8624.2008.01154.x

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly*, *21*(4), 471-490. doi:10.1016/j.ecresq.2006.09.003

McGraw-Hill. (2014). *Reading Wonders.* New York, NY: Author.

Mehta, P. D., Foorman, B. R., Branum-Martin, L., & Taylor, W. P. (2005). Literacy as a unidimensional multilevel construct: Validation, sources of influence, and implications in a longitudinal study in grades 1 to 4. *Scientific Studies of Reading*, *9*(2), 85-116. doi:10.1207/s1532799xssr0902_1

Mplus (Version 7) [Computer software]. Los Angeles, CA: Muthén & Muthén.

NICHD. (2002). The relation of global first-grade classroom environment to structural classroom features and teacher and student behaviors. *The Elementary School Journal, 102*(5), 367-387. doi:10.1086/499709

NICHD. (2005). A day in third grade: A Large-Scale study of classroom quality and teacher and student behavior. *The Elementary School Journal, 105*(3), 305-323. doi:10.1086/428746

Pakarinen, E., Lerkkanen, M., Poikkeus, A., Kiuru, N., Siekkinen, M., Rasku-Puttonen, H., & Nurmi, J. (2010). A validation of the classroom assessment scoring system in Finnish kindergartens. *Early Education and Development*, *21*, 95–124.

Paro, K. M. L., Burchinal, M., Hamre, B. K., Locasale-Crouch, J., Pianta, R. C., Bryant, D., . . . Howes, C. (2009). Quality in kindergarten classrooms: Observational evidence for the need to increase children's learning opportunities in early education classrooms. *Early Education & Development, 20*(4), 657-692. doi:10.1080/10409280802541965

Pearson. (2013). *Reading Street Common Core*. London, UK: Author.

Pianta, R., Howes, C. Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science*, *9*(3), 144-159. doi:10.1207/s1532480xads0903_2

Pianta, R. C., La Paro, K., & Hamre, B. K. (2008). *Classroom Assessment Scoring System*. Baltimore, MD: Brookes.

Pilcher, H. & Kim, Y-S. (2015). *Child by instruction interactions during first grade literacy instruction.* Unpublished manuscript, Department of Curriculum and Instruction, Florida State University and Florida Center for Reading Research, Tallahassee, Florida.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

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

Rudasill, K., Gallagher, K., & White, J. (2010). Temperamental attention and activity, classroom emotional support, and academic achievement in third grade. *Journal of School Psychology*, *48*, 113–134. doi:10.1016/j.jsp.2009.11.002

Rutter, M., & Maughan, B. (2002). School effectiveness findings 1979–2002. *Journal of School Psychology*, *40*(6), 451-475. doi:10.1016/S0022-4405(02)00124-3

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Snijders, T. (2005). Power and Sample Size in Multilevel Linear Models. In B. S. Everitt, & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science*, (Vol. 3), (pp. 1570−1573) Chicester, England: Wiley.

SRA/McGraw-Hill. (2007). *SRA Imagine It!.* Chicago, IL: Author.

Taylor, J., Roehrig, A. D., Soden Hensler, B., Connor, C. M., & Schatschneider, C. (2010). Teacher quality moderates the genetic effects on early reading. *Science*, *328*(5977), 512-514. doi:10.1126/science.1186149

Taylor, B. M., & Pearson, P. D. (2002). Looking inside classrooms: Reflecting on the "how" as well as the "what" in effective reading instruction. *The Reading Teacher*, *56*(3), 270.

Stronge, J. H., Ward, T. J., & Grant, L. W. (2011). What makes good teachers good? A cross-case analysis of the connection between teacher effectiveness and student achievement. *Journal of Teacher Education, 62*(4), 339-355. doi:10.1177/0022487111404241

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63*(3), 249-294. doi:10.2307/1170546

Wechsler, D. (2009). *WIAT III: Wechsler individual achievement test, third edition* . London: The Psychological Corporation.

Wharton-McDonald, R., Pressley, M., & Hampston, J. M. (1998). Literacy instruction in nine first-grade classrooms: Teacher characteristics and student achievement. *The Elementary School Journal, 99*(2), 101-128. doi:10.1086/461918

Woodcock, R. W., McGrew, K., & Mather, N. (2001). *Woodcock-johnson tests of achievement* Itasca, IL: Riverside Publishing.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*(1), 57-67. doi:10.1023/A:1007999204543

BIOGRAPHICAL SKETCH

Heather Pilcher received her Associate in Arts degree from Chipola College in 2007 and her Bachelor's of Science degree in Elementary Education from Florida State University's Panama City Campus in 2009. She taught second grade students for three years and was recognized as her school's Rookie Teacher of the Year in 2010. She received her Master's of Science degree in Elementary Education in 2012 and was then awarded the Pre-doctoral Interdisciplinary Research Training fellowship at the Florida Center for Reading Research at Florida State University. In 2014, she began work as a project manager on the Development of Oral and Silent Reading Fluency research project. During her time in graduate school Heather had the opportunity to present her work at conferences located in Hong Kong, China; Santa Fe, New Mexico; and Kona, Hawaii.