A TWO-STUDY INVESTIGATION OF FIDELITY OF EARLY READING INTERVENTIONS: EXAMINING THE QUALITY OF THE RESEARCH BASE AND AN APPLICATION OF PROGRAM DIFFERENTIATION

A Dissertation

by

MELISSA SHEA FOGARTY

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2012

Major Subject: Educational Psychology

A Two-Study Investigation of Fidelity of Early Reading Interventions: Examining the Quality of the Research Base and an Application of Program Differentiation

A TWO-STUDY INVESTIGATION OF FIDELITY OF EARLY READING INTERVENTIONS: EXAMINING THE QUALITY OF THE RESEARCH BASE AND AN APPLICATION OF PROGRAM DIFFERENTIATION

A Dissertation

by

MELISSA SHEA FOGARTY

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Shanna Hagan-Burke
	Deborah Simmons
Committee Members,	Bruce Thompson
	Erin McTigue
Head of Department,	Victor Willson

August 2012

Major Subject: Educational Psychology

ABSTRACT

A Two-Study Investigation of Fidelity of Early Reading Interventions:

Examining the Quality of the Research Base and an Application

of Program Differentiation. (August 2012)

Melissa Shea Fogarty, B.S., The University of Texas at Austin;

M.Ed., Texas State University

Co-Chairs of Advisory Committee: Dr. Shanna Hagan-Burke
Dr. Deborah Simmons

This research consisted of two studies. The purpose of the first study was to examine the presence and quality of fidelity of implementation as reported in recent early reading intervention research. A comprehensive search of kindergarten through third-grade reading interventions published between the years 2005 and 2011 was conducted. Articles that met the inclusion search criteria were analyzed according to fidelity dimensions. Findings from the first study indicated an increase in fidelity reporting from 2001 to 2005. Few articles, however, analyzed the relationship between fidelity of implementation and student outcomes. While there has been an increase of early reading intervention studies reporting fidelity, there is a lack of studies reporting fidelity in relation to student outcomes. Many studies are reporting multiple dimensions of fidelity, but few studies assess the program differentiation dimension.

The second study was an exploratory study focused on the fidelity dimension of program differentiation as applied to two early reading interventions from an

experimental study. A fidelity observation instrument was created using evidence-based reading practices. The fidelity observation instrument was then used to evaluate instructional practices, teacher responsiveness, and student engagement of an experimental and comparison reading intervention at three time points to examine program differentiation. Latent constructs were created using exploratory factor analysis and were then used to compute an effect size called the achieved relative strength index, which is the difference between two experimental conditions. Findings from the exploratory factor analysis in the second study indicated items loaded onto three latent constructs: (a) instructional practices, (b) teacher responsiveness, and (c) student engagement. The instructional practice achieved relative strength index effect size was large for the experimental group. The achieved relative strength index effect size for both teacher responsiveness and student engagement was small, indicating little difference between the two conditions. The second study in this research endeavor addressed that gap by applying the achieved relative strength index effect size to an early reading intervention study and demonstrating one way to capture program differentiation. Finally, implications for future research were addressed as part of the study.

DEDICATION

My dissertation is dedicated to my beautiful Fogartys. Throughout the entire dissertation process, the three of you kept me going when I didn't want to be going.

Patrick, we did it. We both made it through school. What a crazy, beautiful life we have together! Your encouragement has meant the world to me. I guess now we have to be real grownups with jobs and stuff. There is nobody I would rather be a grownup with than you!

Presley, you are my spark! You are the smartest, most clever person I have ever met. You're the reason I work so hard. I want to teach you that you don't have to be perfect—you just have to work hard. During my grad school journey, I wanted to quit several times. One look at you, and I knew I couldn't. I want you to be as proud of me as I am of you.

Pierce, you are my light! Nothing matters and there is no stress when you sit on my lap and call me mommy. You were born during my second semester of grad school, and I believe you were sent here to get me through this and to remind me of the important things in life.

Even on this day, my sweet P's are my proudest accomplishment. Guess what guys? Mommy is done! No more d-word! I love you three for infinity!

ACKNOWLEDGEMENTS

This was not a solo mission. Many people helped, guided, poked, prodded, and encouraged me along the way.

First, I would like to thank my committee co-chairs, Dr. Hagan-Burke and Dr. Simmons. Dr. Hagan-Burke, I remember driving in to meet you four years ago and we sat at a Starbucks. You were so fun and friendly, and I thought, "How bad could this be?" Thank you for your support, but more importantly, for your encouragement and guidance. I know I have made you batty and have been your problem child, but you have had a huge impact on my graduate life experience. I wish I had a dollar for all the times you said, "You're doing it!" You can now say, "*We* did it!"

Dr. Simmons, thank you for taking me under your wing. I'm overwhelmed with gratitude for the opportunities that you have provided me. I feel like I have grown tremendously from the first ERI meeting. I'm sure there have been some times when I was less than stellar, but with your guidance and example, I have grown in ways I didn't think possible. When completing a task now, I ask myself, "WWDSD," or "What would Dr. Simmons do?" Thanks for being my voice of reason.

Dr. Thompson, you were the Mount Everest that I had to climb. I think I shed a thousand tears over your class and stats. I remember telling you my first semester, "This time last year, I was potty training children in my pre-school class, and now you want me to understand ANOVA?" Fast-forward four years, and now I feel that if I can get through your classes, I can do anything. Thanks for having what felt like impossible

expectations because I learned a lot about who I am by surviving 640, 641, and 642. My graduate life experience would not have been the same without you.

Dr. McTigue, thank you for agreeing to be on my committee. Your friendly presence has always put me at ease. You have been flexible and supportive, and I appreciate that. You are an excellent model on how to be a mommy in academia. Thank you!

Additionally, thank you to the entire EPSY faculty and staff. I feel like we really do learn from the best here at Texas A&M, and that is coming from an orange-bleeding Longhorn.

Of course, I have to mention my cohort and friends, Stacey, Eric, Rose, Leigh Ann, Jaime, and Amy. I'm not sure I would have made it without having other souls who knew and understood exactly what I was going through when I was going through it. Thanks, friends! Stacey, I especially would not have made it without you. I remember locking equally panicked eyes with you in a stats study group when Eric and Rose were explaining something from 640. I knew right then that we were in a secret-friends-for-life pact together. Eric, you have been such a great office buddy. You are always a kind and patient teacher of stats as well. Rose, being mommies together in this process has been so helpful. I always feel like you understand my extra stresses. You are an inspiration with your energy and your momminess. John Davis thanks for getting me into this mess. I'm glad our families are here together.

Finally, thanks to my entire family for their encouragement. My family means the world to me and I'm super proud to be a part of such a wacky, loving clan. Andy,

Parker, and Eric, even though you are the world's best brothers and I am so proud of all three of you, you may now call me Dr. Sis. Thanks! Thank you, Dad, Beth, Mom, and Rob for your ongoing support. It's nice to know that you have my back! I love you all very much!

Grandpa and Grandma, you have taught me that I can do anything I want to do as long as I do it well and with class. Grandpa, I have tried to be the best Georgia buggy worker ever to make you proud. I love you more! Grandma, thank you for always encouraging and checking on me. Your excitement and interest have helped me more than you know. When I grow up, I want to be just like you!

Mom, you have always been such a role model with your drive, energy, and accomplishments. On numerous occasions, I have witnessed your determination. Once you make your mind up that you are going to do something, consider it done; you attack it like a tiger. My whole life people have said that I look and act just like you. Well, thank God for that! I love you so much!

Again, Patrick you have been such a support system. Thanks for all that you do for me and our children. You have supported me through my BS, my MEd, and my PhD! I pinky promise that this is the last time you will ever read over one of my school papers. I pinky promise no more school. EVER!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION TO FIDELITY OF IMPLEMENTATION

Fidelity of implementation refers to "the extent to which an enacted program is consistent with the intended model" (Century, Rudnick, & Freeman, 2010, p. 4). Many terms are used to discuss fidelity of implementation, including treatment integrity, treatment fidelity, and implementation of the independent variable. For the purposes of this study, the term *fidelity* will be used as it pertains to dimensions related to intervention research.

The study of fidelity is critical to causal relationships between a treatment and student outcomes. Fidelity increases our confidence in the effects of an intervention. That is, outcomes of an intervention can only be attributed to the independent variable if fidelity is upheld by the interventionist (O'Donnell, 2008). The constant and continuous documentation of the independent variable is essential for the researcher to be able to make causal statements about the effects.

Over the past decade, educational researchers have been encouraged to measure and report fidelity of implementation more thoroughly (Sanetti & Kratochwill, 2009). In the current Institute of Education Sciences (IES; 2011) Request for Proposal Application, the requirements outlined for future grants mandate that researchers must have a plan to document and measure fidelity of implementation, with specific attention paid to core components of the intervention. The IES also encourages researchers to

---

This dissertation follows the style of *Exceptional Children.*

"describe how fidelity will be incorporated into analysis of the impact of intervention," which makes a "strong applicant" (p. 50).

Recently, a special edition of *School Psychology Review* that was focused on developing a science of treatment integrity (Sanetti & Kratochwill, 2009) documented the importance of fidelity of implementation. Articles in the special issue focused on different aspects of fidelity, such as measurement issues (reliability and validity), different dimensions and theories, and the importance of fidelity reporting. The editors suggested, "It is essential that journal editors and reviewers require authors to relate treatment integrity data to overall outcomes" (Sanetti & Kratochwill, 2009, p. 4). A special issue of *Exceptional Children* proposed quality indicators for group, single-subject, and correlational research to determine evidence-based practices. Gersten et al. (2005) and Horner et al. (2005) identified measuring and reporting fidelity as essential to quality research in both group and single-subject experimental research.

Different dimensions of fidelity of implementation have been proposed. Dane and Schneider (1998) recommended a five-dimensional framework: (a) adherence—the presence or absence of critical components of the intervention; (b) quality—the teacher variables that can make an impact, such as enthusiasm; (c) exposure—the amount of intervention received; (d) student responsiveness—the extent that students are engaged and on task; and (e) program differentiation—the difference between the experimental and control condition.

Others have proposed dimensions that align with and extend how we measure and report fidelity. Gresham (2009) suggested that treatment integrity often encompasses

adherence, competence (or quality of interventionist), and treatment differentiation. Likewise, in the criteria established for special education research, Gersten et al. (2005) considered surface fidelity and quality as important aspects of program implementation. Surface fidelity consists of the key components of an intervention, which should include sufficient time allocated to program and the amount of material covered. Gersten et al. suggested that quality might include measures such as scaffolding procedures, teacher modeling, and corrective feedback. Another set of proposed fidelity dimensions included a structure component and a process component (Mowbray, Holter, Teague, & Bybee, 2003). Structure refers to the framework of the intervention, while process is the way the intervention is delivered. While the various definitions and dimensions have overlapping features, further research is needed to investigate the importance and relation of individual components.

Given the perceived importance of fidelity to causal inferences regarding treatment effects, this dissertation was designed to advance the understanding of treatment fidelity through two related studies. The first involved a comprehensive review and documentation of fidelity reporting in published early reading intervention research articles. The research questions addressed included:

1. To what extent did early reading intervention research studies published between 2005 and 2011 report implementation fidelity? Did fidelity reporting change from 2005 to 2011?
2. What dimensions of fidelity were most commonly reported among early reading intervention studies published between 2005 and 2011?

3. To what extent did the studies examine relations between fidelity of implementation and student outcomes?
4. To what extent did early reading intervention studies published between 2005 and 2011 meet the proposed quality indicators set forth by Gersten et al. (2005)?

The second study focused on the program differentiation dimension of fidelity. The difference between two early reading interventions was observed and assessed based on evidence-based features of effective instruction. The Fidelity Observation Guide (FOG) was created expressly for this study and was used to document instruction in both conditions. Using exploratory factor analysis, data from the FOG collected at three points during the intervention were used to create latent fidelity factors and tested whether experimental and typical practice conditions differed on these dimensions. This study answered the following research questions:

1. Did indicators based on effective teaching practices create latent factors as measured by exploratory factor analysis?
2. Using latent variables of instruction, to what extent did interventions differ as indexed by an achieved relative strength index?

CHAPTER II

EXAMINING FIDELITY OF IMPLEMENTATION IN CURRENT EARLY READING INTERVENTION RESEARCH

Fidelity of implementation is the accuracy and consistency with which an independent variable is executed as the researcher intended (Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000). The purpose of assessing fidelity is to increase confidence that outcomes of an intervention were causally related to the intervention. Failure to measure fidelity can lead to erroneous causal conclusions (Sheridan, Swanger-Gagne, Welch, Kwon, & Garbacz, 2009).

Prior reviews indicate significant variability in the approaches and extent to which fidelity is reported in intervention research. Gresham et al. (2000) reviewed 65 articles published between 1995 and 1999 that investigated interventions involving students with learning disabilities and their reporting of fidelity. Only 12 (18.5%) reported any information on fidelity of implementation. In a 2006 review of literature focusing on children with autism, only 11 of 60 articles (18.3%) operationally defined the independent variable and assessed fidelity (Wheeler, Bagget, Fox, & Blevins, 2006). In a more recent review of 163 articles on behavioral interventions for students with mental retardation published between 1996 and 2006, only 38 studies (36%) reported some type of fidelity (Wheeler et al., 2009).

Over the past decade, there have been numerous calls for educational researchers to measure and report fidelity of implementation more systematically (Sanetti & Kratochwill, 2009). In proposal development guidelines, the Institute of Education

Sciences (2011) emphasizes the need for grant applicants to carefully attend to fidelity of implementation and the way it is assessed, reported, and analyzed. The professional organization Council for Exceptional Children (CEC) and its respective publication outlet, *Exceptional Children,* emphasize the importance of fidelity of implementation in educational research and have defined standards of research and fidelity reporting.

The CEC created a task force to establish quality indicators for special education research and published those standards in a special issue of *Exceptional Children* (Odom et al., 2005). In that special issue, Gersten et al. (2005) described quality indicators for group experimental and quasi-experimental research and articulated essential and desirable features. With respect to fidelity of implementation, Gersten et al. concluded that "information about treatment fidelity is essential in understanding the relationship between an intervention and outcome measures" (p. 157). This fidelity indicator advocates that researchers assess fidelity multiple times throughout the intervention, at minimum use a checklist to capture the key components of the intervention, include some type of interobserver score, and ensure adequate time and intervention coverage was provided to the learner.

Furthermore, the CEC and its Professional Standards & Practice Committee released Classifying the State of Evidence for Special Education Professional Practices: CEC Practice Study Manual (CEC, 2008) to identify criteria needed to determine evidence-based practices. One of the criteria identified as essential for classifying research as an evidence-based practice is fidelity of implementation. Meeting this fidelity criteria requires (a) assessing implementation fidelity throughout the entire

course of the study on a regular basis using a low inference measure; (b) assessing key features of practice using a checklist of critical intervention aspects that are determined to be adequate; (c) when relevant, determining that adequate time was allocated for the intervention; and (d) when relevant, determining that an adequate amount of intervention material was covered.

Fidelity of implementation has also been the focus of the Society for Prevention Research (SPR), an organization devoted to improving prevention research for social, physical, mental health, and academic problems. In 2005, SPR outlined standards of evidence for efficacy and effectiveness trials (Flay et al., 2005). These standards echoed those of Gersten et al. (2005), stating that detailed and precise descriptions of interventions are necessary for replication and arguing that fidelity reporting should be a standard for efficacy trials. The authors (Flay et al., 2005) noted that implementation can vary greatly in efficacy studies implemented under natural conditions and discussed multiple aspects of fidelity including acceptance, compliance, adherence, and/or involvement of the target audience.

Given the importance of fidelity of implementation to interpreting intervention research and the relatively recent standards emphasizing fidelity, this study was designed to examine the state of fidelity in current research. In particular, I was interested in fidelity reporting as related to early reading interventions from kindergarten through third grade. Thus, the following sections summarize the research literature on early reading intervention and dimensions of fidelity reviewed as part of this study.

Early Reading Intervention

Over the past 20 years, a converging evidence base has accrued to support early reading intervention. Primary research and research syntheses report positive outcomes for children who receive early reading intervention (Foorman, Breier, & Fletcher, 2003; Scammaca, Vaughn, Roberts, Wanzek, & Torgesen, 2007). Effective early reading intervention emphasizes multiple dimensions of reading including the content of instruction. Phonemic awareness, decoding, vocabulary, comprehension, and fluency have been identified as necessary foci of instruction (Gersten et al., 2005; National Reading Panel, 2000; Scamacca et al., 2007). In addition, prior research provides evidence of the importance of how intervention is delivered. Critical delivery features of early reading intervention, particularly for students who are at risk for reading difficulties, include explicit and direct instruction, more intensive instruction through small groups and/or more instructional time, and more supportive instruction through scaffolding and feedback (Foorman et al., 2003; Foorman & Torgesen, 2001). Scammaca et al. (2007) synthesized interventions for struggling readers and found that effective reading interventions usually included small or one-on-one group sizes, as well as an almost daily frequency of intervention.

In summary, *what* is taught (content, phonemic awareness, decoding, comprehension, vocabulary, fluency), *how* it is taught (explicit and systematic), *how much* it is taught (intensity, duration, grouping), and the quality of the instruction (opportunities to respond, feedback and scaffolding techniques) are essential elements of early reading intervention that are related to improving the reading trajectories of early

struggling readers. These critical features are particularly relevant for fidelity of implementation for determining what should be measured. Ensuring proper measurement increases confidence that all aspects of the intervention were implemented and assessed, in order to be able to make causal statements and to understand what was implemented for replication efforts of an early reading intervention.

Fidelity of Implementation

The Multiple Dimensions of Fidelity of Implementation

One of the challenges of fidelity of implementation is deciding what to measure. Early reading intervention includes multiple dimensions, and a primary issue in studying fidelity of implementation is what should be measured. Reviews of fidelity of implementation research and guidelines by professional organizations reveal multiple dimensions that can be measured. In 1998, Dane and Schneider proposed a five-dimensional fidelity framework based on a review of public health literature. The dimensions included (a) adherence, (b) exposure, (c) quality of delivery, (d) participant responsiveness, and (e) program differentiation. Dane and Schneider's framework is currently the most frequently cited fidelity model; at the time of this study, there were over 394 citations reported on Google Scholar for this framework. Following is a review of the dimensions of fidelity outlined in Dane and Schneider's framework augmented with related dimensions and definitions identified by other researchers.

Adherence addresses whether the components of the intervention are delivered as intended. Surface fidelity, a term that could be considered synonymous with Dane and Schneider's (1998) adherence dimension, consists of key components of the intervention

being delivered as prescribed (Gersten et al., 2005). In early reading intervention, adherence involves documenting the essential elements as identified by developers or programmers and whether a sufficient proportion of the intervention was implemented.

Exposure, often referred to as dosage, addresses the total amount of intervention received by the participants. Dane and Schneider (1998) articulated three ways that an exposure variable can be collected: (a) "the number of sessions implemented"; (b) "the length of each lesson"; and (c) "the frequency with which program techniques were implemented" (p. 45). Although Gersten et al. (2005) included "adequate time allocation per day or week" and "coverage of specified amount of material in the curriculum" (p. 157) in what they called surface fidelity, those items can be considered examples of the exposure dimension. How much an intervention was implemented can help determine whether an acceptable amount of reading intervention was received by the students and determine the relation between exposure and outcomes. With respect to early reading intervention, exposure may provide important information on to how to intensify intervention to improve student outcomes.

Quality of delivery refers to a qualitative measure of how well the intervention was implemented. Quality can also be referred to as competence (Schulte, Easton, & Parker, 2009). Dane and Schneider (1998) described quality of delivery as a qualitative measure that tries to capture aspects that are not prescribed by the intervention but can have an impact on student outcomes, such as preparedness, enthusiasm, and attitudes. Gersten et al. (2005) discussed the importance of quality in providing insights into the effects or non-effects of an intervention beyond the mere adherence to an intervention's

components. Examples of quality provided by Gersten et al. included scaffolding procedures, teacher modeling, and corrective feedback, which have all been identified as critical to early reading interventions (Foorman et al., 2003; Foorman & Torgesen, 2001). Gresham (2009) concluded, "One can adhere to a particular intervention with perfect integrity yet do so in an incompetent manner" (p. 534).

Participant responsiveness refers to the extent participants are engaged and responsive to intervention. The participant responsiveness dimension helps answer the question of how engaged students were during the intervention. According to Dane and Schneider (1998), this dimension could measure participant enthusiasm and engagement in the intervention. In an early reading intervention, this might measure how much time a student remains on task during the reading session. Theoretically, higher engagement would positively influence reading outcomes.

The program differentiation dimension refers to the difference and comparison in content and instructional practices of all conditions of research. The objective of program differentiation is to determine how alike and different the intervention was from typical practice or from comparison interventions. Dane and Schneider (1998) described this dimension as a "safeguard against the diffusion of treatment" (p. 45). Hulleman and Cordray (2009) emphasized program differentiation in their definition of fidelity as well, stating that the "treatment has to be stronger or different from the counterfactual condition" (p. 91). They described how program differentiation can be calculated as an effect size using dosage (exposure) and participant responsiveness dimensions.

Dane and Schneider (1998) asserted that researchers would have a more complete picture of fidelity by examining all five dimensions and provided a common framework to allow researchers to compare results across studies. Although their framework was designed for health studies, it has particular applicability to early reading intervention and provides a comprehensive net that encompasses dimensions recommended by special education researchers. Using a multiple-dimension approach to fidelity, researchers may be able to better understand what to measure to help increase the reporting of fidelity.

Relating Fidelity of Implementation to Student Outcomes

In addition to a lack of fidelity reporting, studies that actually analyze fidelity as an independent variable are difficult to locate. Very few studies that provide descriptive data on fidelity of implementation of their intervention link those results to student outcomes (Sanetti & Kratochwill, 2009). In fact, in a recent review of K-12 curriculum interventions, O'Donnell (2008) found only five studies that examined the relationship between student outcomes and some type of fidelity measure.

Evaluating the Quality of Fidelity of Implementation

As part of their charge to develop quality indicators for group experimental and quasi-experimental research, Gersten et al. (2005) to created quality indicators to determine evidence-based practices. More specifically for fidelity, Gersten et al. considered the quality indicator "was the fidelity of implementation described and assessed" as an essential indicator for both research proposal and research articles (p.

152). This implies the need to advance the measurement and reporting of fidelity in order for a study to be considered high quality.

Recently, Jitendra, Burgess, and Gajria (2011) and Chard, Ketterlin-Geller, Baker, Doabler, and Apichatabutra (2009) evaluated the literature base of cognitive strategy instruction and repeated readings using Gersten et al.'s (2005) quality indicators to determine the quality of the evidence base. Both of these studies created a rubric to rate the literature on every quality indicator. Because the purpose of this study was focused on fidelity, only the quality indicator relating to fidelity was examined; however, the important work of Jitendra et al. (2011) and Chard et al. (2009) guided this study.

Present Study

Reviews of intervention research largely based on studies conducted prior to 2005 indicated that fidelity reporting was not standard practice. Given the emphasis on fidelity of implementation standards over the past decade, it is reasonable to conclude that more recent research would reflect a higher quantity and quality of fidelity reporting. This study was particularly focused on the state of fidelity of implementation measurement and reporting in early reading intervention research published since 2005.

The purpose of this study was to examine the state of fidelity of implementation in the body of experimental early reading intervention group studies published between 2005 and 2011. This period of research was examined because it reflects the period since criteria for fidelity reporting were published by the CEC (Gersten et al., 2005). To provide a comprehensive lens by which to study fidelity dimensions, this study adapted

Dane and Schneider's (1998) framework to determine the range and extent to which their dimensions were reflected in the early reading research. To examine the quality of fidelity reporting, the study operationalized criteria outlined by Gersten et al. (2005). The investigation focused on experimental early reading intervention studies involving students in kindergarten through third grade. The following questions guided this study.

1. To what extent did early reading intervention research studies published between 2005 and 2011 report implementation fidelity? Did fidelity reporting change from 2005 to 2011?
2. What dimensions of fidelity were most commonly reported among early reading intervention studies published between 2005 and 2011?
3. To what extent did the studies examine relations between fidelity of implementation and student outcomes?
4. To what extent did early reading intervention studies published between 2005 and 2011 meet the proposed quality indicators for fidelity set forth by Gersten et al. (2005)?

Method

Literature Search

A comprehensive search of early reading intervention studies published between 2005 and 2011 was conducted within the PsycInfo, ERIC, and Education Full Text (Wilson) databases. Key search terms included *reading, reading intervention, early literacy intervention, elementary, comprehension, fluency, decoding, phonemic awareness*, and *vocabulary*. In addition, the reference list for each selected article was

searched using the Scopus database. Finally, a hand-search of the current issues of the following journals was conducted: *Exceptional Children, Reading Research Quarterly, Journal of Special Education, Educational Researcher, Elementary School Journal, School Psychology Review, Journal of Educational Psychology, Journal of Educational Research, Journal of Learning Disabilities, Learning Disabilities Research and Practice, Journal of Literacy Research, Remedial and Special Education, Scientific Studies of Reading,* and *Journal of Research on Reading.*

Selection Criteria

Each article was evaluated using the following criteria. First, studies had to employ a group or quasi-experimental design published between 2005 and 2011 in a peer-reviewed journal in the United States and printed in English. These dates were chosen because the quality indicators set forth in *Exceptional Children* were published in 2005, the year in which increasing attention was placed on fidelity (Flay et al., 2005; Gersten et al., 2005; Horner et al., 2005). While studies published in 2005 were not likely to reflect the new quality indicators, they were included to serve as a baseline.

Studies included participants in kindergarten through third grade. These grades were chosen primarily for their focus on early reading. Studies of students in grades higher than third were excluded even if a portion of the sample was in kindergarten through third grade. A broad range of child participants was represented in the selected studies, including English language learners as long as they were receiving reading intervention delivered in English.

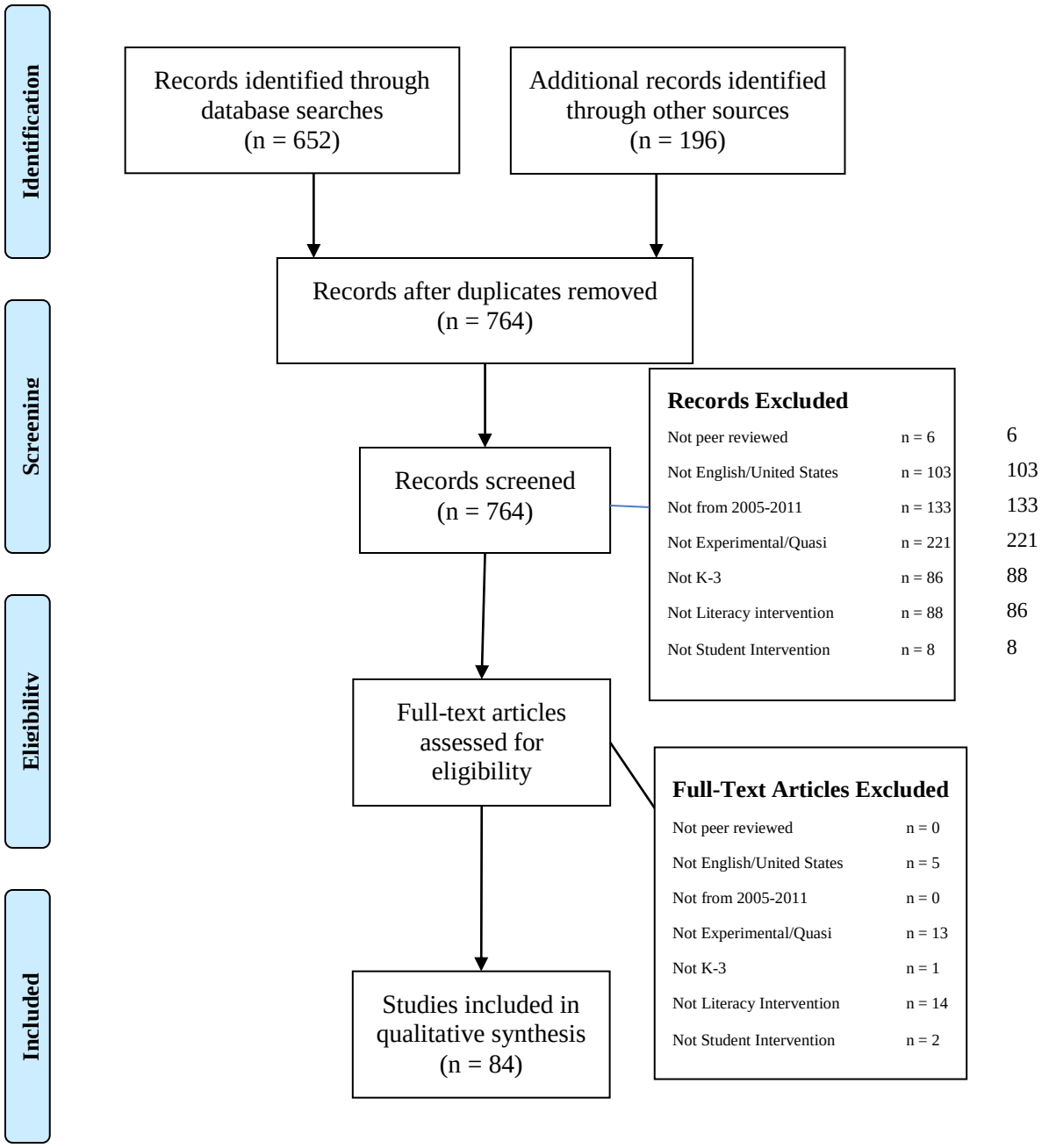


Figure 2.1. Flow diagram of articles selected.

The research in the selected studies had to specifically address one of the National Reading Panel's (2000) suggested reading components for grades K-3

(phonemic awareness, decoding, vocabulary, comprehension, and fluency). Studies with interventions that solely addressed writing were excluded. Finally, only studies that focused on intervention with students (as opposed to teachers or parents) were included. Studies that included peer tutors were also selected. Figure 2.1 provides a visual summary of the search and selection process using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Guide (PRISMA; Liberati et al., 2009).

A total of 778 articles were screened using the selection criteria previously described. To examine the reliability of the screening process, a subset of 195 articles (25%) was randomly selected from the original 778 and another coder independently screened each using the same selection criteria. Interrater agreement was calculated as percent of agreement by dividing agreements by agreements plus disagreements, and multiplying by 100. The initial overall percent agreement for studies included was 83.9%. All of the disagreements regarded the design of the study. For some of the articles, coders had difficulty discerning whether the design was experimental or quasi-experimental. Coding discrepancies were resolved by a third rater.

Coding Procedures

A database was developed using the online software service Zoho Creator to systematically code selected variables from each of the included articles. Nine overall fidelity variables and six quality indicators were coded for each of the 84 articles selected. Coders were graduate students in special education who received a minimum of 2 hours of individualized training. The Zoho service optimized coding decisions and subsequent interrater reliability checks by providing drop-down boxes and checklists to

force choices. Each of the 84 articles was independently coded twice. There were few disagreements, and the overall percentage of agreement rate was 95%. In the few instances where there was a discrepancy, the problem was reviewed by the first author, who determined what the correct code should have been, and then discussed with the coders to clarify any misunderstandings and prevent future discrepancies.

Fidelity variables. Each article was coded to capture comprehensive information regarding fidelity as reported in the current early reading intervention research. Figure 2.2 provides a list of the fidelity variables coded as well as their descriptions from the codebook. Nine variables were coded from two different categories: essential fidelity variables and measurement of fidelity variables. The essential fidelity variables included the following items: (a) Was fidelity reported? (b) Was fidelity summarized in a quantifiable way? (c) Was fidelity score used as an independent variable? (d) Select all of the dimensions of fidelity that were reported. The measurement of fidelity variables included the following: (a) How was fidelity coded (live observation, video, etc.)? (b) How many times was fidelity assessed? (c) How was fidelity measured (checklist, rating scale, etc.)? (d) How was interrater reliability reported?

Essential Fidelity Variables

1. **Was fidelity reported?**
   Select yes or no. (If the word fidelity is mentioned, select yes.)
2. **Was fidelity summarized in a quantifiable way (descriptive, percentages, etc.)?**
   Select yes or no. Select yes, if the author presented any number values for fidelity. (Example, authors stated that teachers had an average fidelity of 92%.)
3. **Was a fidelity score used as an independent variable (i.e., moderator or linked with student outcomes)?**
   Select yes or no. (Select yes, if any one of the fidelity dimensions was used as a predictor of student outcomes.)
4. **Select all of the dimensions of fidelity that were reported.**
   - **Adherence:** *Were the components of the intervention delivered, and to what extent?*
   - **Quality:** *Did the researchers determine how well the intervention was performed? Was teacher quality taken into consideration? (Example items include pacing, scaffolding, readiness, etc.)*
   - **Program Differentiation**: *Did the researchers measure similar components in both conditions to be able to compare numerically?*
   - **Student Responsiveness**: *Did research team measure how engaged or on task the students were? Was there an item that related to how well or how often the students responded to intervention?*
   - **Dosage/Exposure**: *What is the total amount of intervention that the student received? This could be measured in minutes or days. The researcher must refer to dosage.*

Measurement of Fidelity Variables

5. **How was fidelity coded?**
   Select live observation, video, audio, permanent product, teacher self-report, student report, not reported, or other.
6. **How many times was fidelity assessed?**
   Write in the number of times that fidelity was assessed as reported by the authors.
7. **How was fidelity measured?**
   Select checklist, rating scale, field notes, or other.
8. **How was interrater reliability reported?**
   Select Kappa, percent agreement, other, or none reported.

Figure 2.2. Fidelity variables for coding articles.

Quality indicator codes. The quality indicator rubric created for this study, was based on Gersten et al.'s (2005) quality indicators for group research and modeled after

rubrics published by Jitendra et al. (2011) and Chard et al. (2009), in which indicators were used to determine the quality of the evidence for cognitive strategy instruction and repeated readings. In the present study, the rubric was designed specifically for fidelity; therefore, the broader research quality indicators published by Gersten et al. (2005) were not included.

The fidelity rubric was created using terminology from the Gersten et al. (2005) article, which suggests essential and desirable fidelity indicators. The fidelity rubric addresses the following question: "Are the procedures for ensuring and assessing fidelity of implementation described?" (Gersten et al., 2005, p. 151). For the purposes of the fidelity rubric, this question was disaggregated into indicators to include more detailed information as provided in the Gersten et al. article. The indicators scored for coding were as follows: (a) was the data collection method described? (b) was interrater reliability reported? (c) was the fidelity measure described? (d) was fidelity monitored throughout the intervention? and (e) was fidelity data analysis reported? Scoring employed a 3-point rating scale (0 = indicator not met, 1 = indicator partially met, 2 = indicator met with high quality).

Results

Percent of Studies Reporting Fidelity

Of the 84 early reading intervention studies examined in this study, 60 (71.4%) reported some type of fidelity. This variable was coded for each study as yes or no. Any reference to fidelity was coded as yes without regard to the extent to which it was measured. Data were examined by year to determine whether fidelity reporting increased

after the quality indicators were published in 2005. Figure 2.3 illustrates the change in fidelity reporting over the years. In 2005, nine of 16 articles (56.2%) reported fidelity. In 2006 and 2007, the percentage of articles reporting fidelity increased to 75.0%. For 2008 and 2009, the percentage decreased slightly to 66.7%. By 2011, the percentage of articles reporting fidelity increased to 91.7%. A *t*-test indicated a statistically significant difference (2.16, $p < .041$) between the years 2005 ($M = .56$, $SD = .5$) and 2011 ($M = .92$, $SD = .28$), underscoring a 35.4% increase in the number of early reading intervention studies reporting fidelity during the 6-year period.

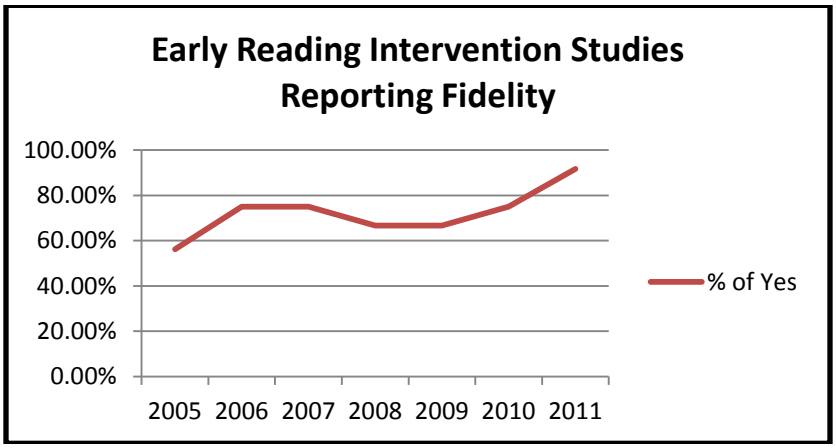


Figure 2.3. Percentage of published early reading intervention studies reporting fidelity by year.

Fidelity Dimensions Reported

Adherence. The most common fidelity dimension reported was adherence, with 58 (69%) studies referring to some type of fidelity. Adherence was the sole dimension of fidelity reported in 27 (32.1%) of the studies, while an additional 31 (36.9%) studies

reported adherence in combination with other dimensions. Across the studies reviewed, adherence was defined and discussed in a variety of ways. Synonymous terms for adherence included integrity checks and procedural fidelity (Kamps et al., 2007; Simmons et al., 2011). Some of the studies measured adherence by measuring the presence or absence of components (e.g., Fien et al., 2011; Puhalla, 2010). Studies by Coyne, McCoach, and Kapp (2007) and Coyne, McCoach, Loftus, Zipoli, and Kapp (2009) developed a checklist based on the different components of the intervention, while Zipoli, Coyne, and McCoach (2011) created an adherence instrument based on a task analysis of important instructional steps.

Quality. Twenty-two (26.2%) of the studies included some measure of quality. In each of these instances, quality was reported along with one or more dimensions of fidelity. Across the studies, quality encompassed multiple features including how well the intervention was performed, effective teaching practices, and overall or global quality. Mathes et al. (2005) evaluated quality as a global checklist of (a) "readiness of instructional materials," (b) "appropriate student seating arrangement," and (c) "instructor warmth and enthusiasm" (p. 159).

With respect to effective teaching, Gunn, Smolkowski, and Vadasy (2011) included four quality components referred to as teacher-student instructional interactions related to instruction: (a) "teacher demonstration or model," (b) "an independent student practice," (c) "a student error," and (d) "corrective feedback" (p. 63). Vadasy and Sanders (2008b) incorporated the following instructional quality behaviors: (a) maximizes time on instruction, (b) uses quick pace/smooth transition, (c) offers

appropriate specific praise, (d) organizes materials, (e) maintains accurate attendance records, and (f) provides appropriate error correction. Using a Likert scale, Frechtling, Zhang, and Silverstein (2006) measured "the extent and quality to which teachers" (a) provided small group instruction, (b) used the curriculum guide, (c) used reading stations, and (d) provided interventions for struggling students (p. 89).

Other approaches for measuring quality were used by Wanzek and Vaughm (2008) and Vaughn et al. (2009), who calculated a quality of instruction composite score that reflected the occurrence of intervention components, appropriateness of instruction and material, and rating for instructional time. Simmons et al. (2011) reported a global instructional quality score for the entire lesson.

Dosage. Seventeen (20.2%) studies reported some measure of dosage. The Amendum, Vernon-Feagans, and Ginsberg (2011) study described dosage as duration of intervention, expressed as the total number of weeks students received intervention as reported on a 5-point Likert scale (1 = no weeks of intervention, 3 = 4 to 9 weeks of intervention, 5 = 19 weeks or more of intervention). Other examples of dosage included time spent in instruction calculated as time allocated to each intervention component as well as the entire instructional time (Simmons et al., 2007, p. 337). Silverman (2007) reported average dosage amounts ranging from 24 to 35 minutes per teacher.

Student responsiveness. Twelve (14.3%) of the studies reported student responsiveness as part of their fidelity measures. If a fidelity measure included one or more items regarding student engagement and/or problem behaviors, it was coded as addressing responsiveness. For example, student responsiveness was captured in the

general teaching behavior observations by Nelson, Vadasy, and Sanders (2011) with a 4-point rating scale to determine "whether or not all children were responsive during the session" (p. 195).

Maynard, Pullen, and Coyne's (2010) fidelity form included six items related to student responsiveness. Items included stems such as the following: (a) students appear to be engaged in the introduction of target words, (b) students appear to be engaged in the storybook reading, and (c) students maintain appropriate behavior throughout lesson. Their fidelity form also included an area for the observer to tally task behavior.

Program differentiation. Ten (11.9%) studies included fidelity measures to capture the difference between treatment and comparison groups. The two general approaches for describing program differentiation were statistical and descriptive. For example, Mathes et al. (2005) tested two conditions to see if there was a statistically significant difference on measures such as pacing, scaffolding, maintaining students' attention, and following procedures. Pacing was the only item found to have a better effect in the treatment group than in the comparison group ($t(252) = 3.39$, $p < .001$). Similarly, Silverman (2007) conducted a test comparing time spent on read-alouds for both treatment and control and found no statistically significant differences. Alternately, Kamps et al. (2007) descriptively compared both conditions on duration devoted to reading instruction and types of reading done in both conditions based on percentages gleaned from observations.

Multiple fidelity dimensions. As Appendix A illustrates, most studies reported more than one dimension of fidelity. Most studies reported the single dimension of

adherence (32.1%), while 14 (16.6%) studies reported two dimensions. Within the two-dimension category, adherence and quality were the most common combination. An example of this was the Wanzek and Vaughn (2008) study that included both an "implementation validity checklist" or adherence and a "quality of instruction" score.

Thirteen (15.4%) studies reported a combination of three dimensions. Adherence, quality, and student participation were the most common dimensions measured in combination. Wang and Algozzine (2008) reported these three dimensions. Fidelity as an adherence measure was evaluated using a rating scale for each activity to ensure the intervention was implemented as intended. Within the rating scale, the student responsiveness dimension was captured by including an item that measured whether the students were engaged and attentive. Finally, a global checklist with items such as instructor warmth, instructor enthusiasm, and readiness of materials was measured as an overall score. This item was coded as a quality score for the present article.

Five (5.9%) studies reported four dimensions of fidelity. The most prevalent combination was adherence, program differentiation, quality, and student responsiveness. McMaster, Kung, Han, and Cao (2008) used a Kindergarten Peer-Assisted Literacy Strategies (K-PALS) checklist to assess interventionist fidelity to the intervention. The team also used the Classroom Atmosphere Rating Scale (CARS; Wehby et al., 1993) to rate both treatment and control teachers on quality and student responsiveness. They used this information to test for differences, or program differentiation, between the two conditions. McMaster et al. reported that both conditions received moderate to high scores on the CARS rating scale and that there

were no significant differences between the treatment and control groups on the quality and student responsiveness dimension.

Only one article reported all five dimensions (Denton et al., 2010). In a study by Denton et al. (2010), the terms *adherence* and *quality* were used to describe the measures, but example items were not provided. Student responsiveness was measured by students' on-task behavior during intervention. The research team also observed both treatment and control groups on the Instructional Content Emphasis-Revised (ICE-R; Edmunds & Briggs, 2003) and compared "instructional emphasis" for program differentiation. The treatment and control groups were compared using percentages for each content area coded. For example, the treatment condition spent 30.4% of time allocated to fluency, while the typical school practice had 13.5% of time allocated to fluency. Dosage was computed using teachers' attendance records.

Percent of Studies Linking Fidelity to Student Outcomes

Only seven of 84 (8.6%) studies linked fidelity to student outcomes. Of the seven, all but one found a statistically significant relation between a dimension of fidelity and outcomes (Simmons et al., 2007). Table 2.1 displays the results along with the dimensions reported.

Table 2.1

Early Reading Studies That Examined the Effects of Fidelity on Student Outcomes

	Dimensions of Fidelity Examined in Relation to Student Outcomes				
Article	Adherence	Dosage	Program Diff.	Quality	Student Part.
Gunn et al. (2011)	No[1]	Yes[2]	---[3]	No	---
Nelson et al. (2011)	Yes	n/a	n/a	Yes	---
Vadasy et al. (2006)	No	Yes	---	---	---
Ehri et al. (2007)	Yes	---	---	---	---
Frechtling et al. (2006)	Yes	---	---	---	---
Simmons et al. (2007)	No	---	---	---	---
Vadasy & Sanders (2009)	Yes	---	---	---	---

[1] No = Effects on student outcomes were not statistically significant.
[2] Yes = Effects on student outcomes were statistically significant.
[3] --- Indicates that dimension was not examined in relation to student outcomes.

Dosage was analyzed in two of the articles with findings indicating statically significant effects (Gunn et al., 2011; Vadasy, Sanders, & Peyton, 2006a). Gunn et al. (2011) calculated dosage as units complete, with results per teacher ranging from 5 of 26 units to 22.5 out of 26 units. Dosage was associated with student outcomes on several measures such as letter sounds ($r = .31$, $p = .03$), sight words ($r = .71$, $p = .0001$), Comprehensive Test of Phonological Processing (CTOPP; $r = .41$, $p = .0054$), and Peabody Picture Vocabulary Test (PPVT; $r = .34$, $p = .02$). Likewise, Vadasy et al. (2006a) found dosage, calculated as both number of intervention hours and lesson coverage, respectively, to be a statistically significant predictor of most posttest measures, such as alphabetic principle ($r = .40$, $p < .05$; $r = .47$, $p < .05$); reading

accuracy ($r = .42$, $p < .01$; $r = .50$, $p < .01$); comprehension ($r = .40$, $p < .05$; $r = .36$, $p < .05$); oral reading rate ($r = .41$, $p < .05$; $r = .41$, $p < .05$); and spelling ($r = .39$, $p < .05$; $r = .55$, $p < .01$).

The adherence dimension produced mixed results. Three articles found no statically significant effects of adherence to outcomes (Gunn et al., 2011; Simmons et al., 2007; Vadasy et al., 2006a). Others found small correlations between adherence and student outcomes (Ehri, Dreyer, Flugman, & Gross, 2007; Nelson et al., 2011). Ehri et al. (2007) reported very little relationship between fidelity and student outcomes, with r values ranging from .06 to .20 ($p < .05$). Nelson et al. (2011) reported comparable outcomes, with $r = .22$ ($p < .05$).

Two articles found substantial effects when using fidelity, particularly adherence scores, as an independent variable. Frechtling et al. (2006) used adherence scores as ordinal variables to create three categories of implementers: high, medium, and inadequate. Students with teachers that were high implementers outperformed students in both medium and inadequate implementers' classrooms. Students in high-implementer classrooms gained an average of 6.99 points above the medium group and 13.98 above the inadequate group on the letter naming fluency subtest of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002). Vadasy and Sanders (2009) reported that fidelity uniquely predicted word reading and fluency as measured by the Woodcock Reading Mastery Test (WRMT) and Oral Reading Fluency. They found that with all other variables held constant, one standard deviation higher than the mean on tutoring fidelity predicted a 2.77 point gain for word reading and an

11.31 and 13.87 words correct per minute gain on two different passages in reading fluency assessments.

Quality Indicators

The quality indicator rubric created for this study, a fidelity rubric (see Table 2.2), was based on the Gersten et al. (2005) research quality indicators and addressed the following question: "Are the procedures for ensuring and assessing fidelity of implementation described?" (p. 151). Scoring employed a 3-point rating scale (0 = indicator not met, 1 = indicator partially met, 2 = indicator met with high quality). Earning a 2 on an indicator signified that the article presented all information necessary to be considered high quality. Earning a 1 signified that the article presented limited or partial information and should have provided more to be considered high quality. Earning a 0 signified that no information was present in the article regarding the indicator and therefore the indicator was not met.

Table 2.2 displays the means for each quality indictor item as well as a description of coding procedures and includes the (a) data collection method described (M = 1.02, *SD* = .79); (b) interrater reliability reported (M = .44, *SD* = .70); (c) fidelity measurement described (M = .86, *SD* = .77); (d) fidelity monitored (M = 1.09, *SD* = .85); and (e) fidelity analysis reported (M = .71, *SD* = .63). Two of the indicators were partially met, and three of the indicators were not met at all.

Table 2.2

Mean Score of the 84 Articles Using the Fidelity Rubric

Indicator	Mean (*SD*)	Score Criteria		
		Not Met 0	**Partially Met 1**	**Met with High Quality 2**
		Was the fidelity of implementation described and assessed?*		
Data collection method described	1.02 *(.79)*	Provided no information on data collection	Provided limited information	Provided specific information on how many times, ways of collecting data (direct observation, permanent product, survey)
Interrater reliability reported	.44 *(.70)*	None reported	Reliability data did not meet minimal standards or 20% of data not double coded	Reported reliability data with a minimal standard of IOA = 80% or Kappa = 60% and at least 20% of data was double coded
Fidelity measurement described	.86 *(.77)*	Fidelity measures not reported	Fidelity measure briefly described	Provided specific fidelity measure information including dimensions, Likert scale vs. checklist, items on measure, and scoring information
Fidelity monitored	1.09 *(.85)*	No monitoring reported	Fidelity monitored once or twice	Monitored fidelity multiple times throughout intervention period
Fidelity analysis reported	.71 (.63)	No fidelity data reported	Descriptive of fidelity reported	Reported fidelity as an independent variable or as a moderator

*Taken directly from Gersten et al. (2005).

Fidelity monitoring. The highest score was in fidelity monitoring (M = 1.09, *SD* = .85). Scores suggested, as a whole, that fidelity was monitored at least once or twice throughout an intervention. Twenty-seven studies scored a zero (32.1%) or no information on fidelity monitoring reported. Twenty-two studies (26.2%) scored a 1,

indicating that fidelity was monitored once or twice. Thirty-five articles (41.1%) scored a 2, indicating fidelity was measured multiple times throughout the intervention period.

No standard method for monitoring fidelity was identified across the studies. Some articles reported how many total observations were conducted but did not specify how many observations per interventionist. Some articles reported times generally, such as once a month, weekly, or quarterly. For the studies that specified the number of times that fidelity was monitored, the range was from one to 26. Some studies reported percentages of times monitored, and those percentages ranged from 10% to 56%.

Data collection. The quality indicator score for the data collection method was M = 1.02, *SD* = .79. Twenty-five articles scored a zero (29.8%) or no fidelity data collection described. Thirty-two articles (38.1%) scored a 1, indicating that limited information was provided as to how fidelity data were collected. Twenty-seven articles (32.1%) scored a 2, indicating that specific information about data collection procedures were reported, such as number of times and ways of collecting data (direct observation, permanent product, survey).

The majority of studies (*n* = 55, or 91%) that reported fidelity data collection methods used direct observation. Two studies used audio and video tapes of the intervention (Begeny et al., 2010; Fuchs, Compton, Fuchs, Bryant, & Davis, 2008). Another method of collecting fidelity measures was through teacher self-reporting. Four of the 84 articles that reported fidelity chose this method. Nelson, Stage, Epstein, and Pierce (2005) had tutors complete a 17-item self-evaluation using a 4-point rating scale ranging from *never* to *always*.

Description of fidelity measurement. The quality indicator score for fidelity measurement described was M = .86, *SD* = .77. Thirty-one articles scored a zero (36.9%) for no fidelity measurement described. Thirty-three articles (39.3%) scored a 1, indicating that fidelity was briefly described. Twenty articles (23.8%) scored a 2, indicating that specific fidelity information was included, such as rating scales, checklists, items on measurement, and scoring information.

Table 2.3 outlines the types of fidelity measurement reported in early reading intervention studies and their frequencies among the 84 articles coded. After the "did not report" category (39.3%), the rating scale was most common, with 21 out of 84 (25.0%) studies using some type of rating scale. The rating scales ranged from 3 to 5 points. Nelson et al. (2011) used a 5-point scale with the indicators ranging from 0 to 4, with 0 meaning *never* and 4 meaning *proficient* for the criteria. Similarly, Kuhn et al. (2006) used a 5-point scale but rated 1 as *no fidelity* and 5 as *very high fidelity*. Frechtling et al. (2006) used a 4-point scale, ranging from 0 (*not implementing*) to 3 (*implementing effectively*).

Another popular way to measure fidelity was with a checklist. The checklist usually contained a dichotomous choice of "yes" or "no" for items that were relevant to the intervention. Of the 84 studies, 17 (or 20.2%) used a checklist alone, and seven used a checklist in conjunction with other methods. Four of the studies investigating peer-assisted learning strategies (PALS) used the PALS fidelity 40-item checklist to capture fidelity (Calhoon, Al Otaiba, Greenberg, King, & Avalos, 2006; McMaster, Fuchs, Fuchs, & Compton, 2005; McMaster et al., 2008; Rafdal, McMaster, McConnell, Fuchs,

& Fuchs, 2011). The most extensive checklist contained 79 items (Fuchs et al., 2008). Others had a more simplified approach, such as an eight-step treatment fidelity checklist with items such as "following script, modeling, using manipulatives, feedback" (Lo, Wang, & Haskell, 2009, p. 18).

Table 2.3

Types of Fidelity Measurement Used in Early Reading Intervention Studies and Their Frequency

Type of Measurement	Total	%
Did Not Report	33	39.3
Rating Scale	21	25
Checklist	17	20.2
Checklist and Rating Scale	5	6.0
Field Notes	3	3.6
Other	1	1.2
Checklist, Field Notes	1	1.2
Checklist, Field Notes, Rating Scale	1	1.2
Rating Scale, Field Notes	2	2.4
Total	84	100.1

Most articles used a fidelity measure created specifically for the intervention being investigated; however, nine used a published observation protocol as a fidelity measure in some way. McMaster et al. (2008) employed CARS (Wehby et al., 1993) to determine the quality of instructional environment. Two studies used a modified version of the Center for the Improvement of Early Reading Achievement (CIERA) School Change Classroom Observation Scheme (Taylor & Pearson, 2000) to calculate program

fidelity to fluency-oriented reading instruction (FORI; Stahl & Heubach, 2005) and fluency intervention (Kuhn et al., 2006; Schwanenflugel et al., 2009). The ICE-R (Edmunds & Briggs, 2003) was utilized in four studies to examine content and activities that took place in both the experimental and control conditions (Denton et al., 2010; Vadasy & Sanders, 2008b; Vadasy & Sanders, 2010; Vadasy & Sanders, 2011).

Fidelity analysis reported. The fidelity data analysis reported score was M = .71, *SD* = .63. Thirty-two articles scored a zero (38.1%), as no fidelity data were reported. Forty-five articles (53.6%) scored a 1, indicating that fidelity information was described using descriptives. Consistent with previous results, only seven articles (8.3%) scored a 2, indicating fidelity score was used as an independent variable or moderator.

Interrater reliability reported. The lowest quality indicator score was for interrater reliability reporting (M = .44, *SD* = .70), indicating the quality indicator was not met. Fifty-seven articles scored a zero (67.9%) for no interrater reliability reported. Seventeen articles (20.2%) scored a 1, indicating that interrater reliability reporting did not meet minimal standards. Ten articles (11.9%) scored a 2, indicating that reliability information met standards of at least 20% of data being double coded and a percent agreement of 80% or a Kappa score of 60%.

Of the 84 studies, 57 (67.9%) did not report any interrater reliability index. The other articles reported percent agreement, Cohen's Kappa, Pearson r correlations, and intraclass correlation coefficients. Of the 84 articles, 11 (or 13%) reported percent agreement, and five of the 84 (or 6%) calculated Cohen's Kappa. The "Other" category accounted for nine of 84 (or 10%) of the studies, with Pearson r correlations the most

frequently used method. Three studies scored a 2 on every category, for a mean score of 2 (Ehri et al., 2007; Gunn et al., 2011; Vadasy & Sanders, 2009). The three studies are displayed in Table 2.4, along with quality indicator scores.

Table 2.4

Exemplar Articles With a Score of 2.0 out of 2.0

Authors and Score	Data Collection Method Described	Interrater Reliability Reported	Fidelity Measure Described	Fidelity Monitored Multiple Times	Fidelity Data Analysis Reported
Vadasy & Sanders (2009) Score 2	Live observations by four research observers.	Prior to visits, each observer viewed video tapes. Cronbach's alpha of .96, .88, and .97 was computed.	5-point Likert scale coding for adherence and instructional behaviors (items listed).	258 observations averaging 18.	Adherence = 4.5, instruction behavior = 4.6. HLM analysis found that tutor fidelity predicts a 2.77 point gain for word reading and a 11.31 WCPM gain on PRF-U and 13.87 WCPM on PRF-A.
Gunn et al. (2011) Score 2	Observations by trained project staff for entire literacy period.	Four stage process: 1. Overview of system and codes. 2. Practiced on videos. 3. Practiced live with trainer. 4. Met 80% standard. After collecting observation data, an ICC of .85.	3-point rating of fidelity for each activity; global rating on the same 3-point scale (1-low, 2-medium, 3-high).	113 observations for 26 teachers in treatment; 122 observations for 28 teachers in control condition.	Found no significant outcomes for mean activities or global fidelity scores; dosage, or units completed significantly predicted several end of kindergarten outcomes.
Ehri et al. (2007)	Tutors kept logs of components implemented. Two professionals scored for adherence.	Two judges rated all of the records independently and had a 90% agreement.	Rated the tutoring records on 4-point scale; listed all items that were scored along with criteria used to rate.	All records were examined.	4-point scale. Adherence mean for tutors was 13 out of 24. Correlations with posttest reveled little relationship, with values ranging from r = .06 to .20 (p > .05).

Discussion

An analysis of 84 early reading intervention articles published between 2005 and 2011 yielded important information about the current state of fidelity reporting. Findings indicated that researchers are reporting fidelity of implementation more frequently. Fidelity reporting was at its highest in 2011, with 91.7% of early reading studies reporting some type of fidelity. This represents a 35% increase in fidelity reporting from the year 2005.

Dimensions of Fidelity

Of the 84 articles reviewed for this study, adherence was the most common dimension of fidelity reported. It was also the only dimension that was reported independently (i.e., as a standalone fidelity measure with no other dimensions). When any of the other dimensions were reported, they were in combination with at least one other dimension. Increasingly, researchers are measuring more than the simple presence or absence of intervention components. Measures reflecting how students respond to the intervention, how much of the intervention students receive, the quality of instruction during intervention, and program differentiation are all being considered and reported more frequently since the publication of the CEC's quality indicators in 2005 (Gersten et al., 2005).

Dane and Schneider (1998) argued that multiple dimensions are necessary to be able to determine the true nature of an intervention and its effects. Some of those dimensions (i.e., student responsiveness) may be moving beyond fidelity as a construct to reflect intervention outcomes (Schulte et al., 2009). For example, if a teacher employs

an intervention and it increases students' engagement, should student engagement be regarded as an outcome variable or as a fidelity variable? While it is important to capture enough information to accurately explain the effects or lack of effects of an intervention, fidelity as a five-dimension construct may not always be necessary. For example, measuring adherence and quality of implementation may be unnecessary for interventions delivered via computer software. A computer automatically implements with the same procedures and quality for each student. However, dosage, student responsiveness, and program differentiation would be important considerations when deliberating intervention aspects responsible for producing effects. In contrast, adherence, quality, dosage, and student responsiveness would be of particular importance if inadequate response to an intervention increases the students' level of intervention or triggers special education placement, as in a response to intervention (RTI) framework (Schulte et al., 2009).

Fidelity and Outcomes

Overall, only seven of the 84 early reading studies published findings that analyzed fidelity as it relates to student outcomes. This is consistent with the oft-cited O'Donnell (2008) fidelity review, in which O'Donnell looked at K-12 intervention studies in an attempt to identify studies that examined fidelity and student outcomes; only five such articles were found. All five studies reported statistically significant outcomes with higher fidelity. However, the seven articles in the present study that examined fidelity in relation to student outcomes have more inconsistent results. Dosage proved to be a promising predictor, as the studies that examined dosage reported that

students who were exposed to more intervention outperformed their peers who received less amounts of intervention. This may seem rather obvious, but it has practical implications. If teachers can be shown data correlating their usage of an intervention with positive student outcomes, it may increase their willingness and diligence to increase dosage. This information might also motivate teachers to adjust their instructional pace, especially if it is too slow, to increase dosage.

The adherence findings, on the other hand, were more divided. Two studies found that adherence predicted student outcomes (Frechtling et al., 2006; Vadasy & Sanders, 2009). Frechtling et al. (2006) used cut scores and ordinal data coding and found that students in classrooms with teachers who implemented with high fidelity outperformed students with teachers who implemented with medium or low fidelity. Vadasy and Sanders (2011) found that fidelity, with every other variable held constant, could predict 13.87 words read correct per minute on a reading passage.

Five other studies reported little or no relation between fidelity and student outcomes (Ehri et al., 2007; Gunn et al., 2011; Nelson et al., 2011; Simmons et al., 2007; Vadasy et al., 2006). Upon closer examination, however, each of those articles reported high fidelity with low variability. In fact, the authors of one study commented, "The high average of fidelity and limited variance in tutor quality limited its use as a predictor" (Vadasy, Sanders, & Peyton, 2006b, p. 369). This may lead to the conclusion that when high fidelity is achieved and there is little variation among interventionists, researchers have enough information to state causal claims. However, this should be done with caution and only with reliable and valid data sources.

Quality Indicators

One of the goals of this study was to assess the quality of fidelity reporting among early reading studies published between 2005 and 2011 using Gersten et al.'s (2005) "Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education" as a guide. A rubric was created specifically for fidelity reporting, and the indicators were scored on a 3-point scale (0 = indicator not met, 1 = indicator met, 2 = indicator met with high quality). The indicators included data collection description, description of measurement, fidelity monitoring, fidelity analysis, and interrater reliability reporting. As a whole, the early reading intervention studies partially met the quality indicators for data collection and fidelity monitoring. The quality indicators were not adequately met among the studies we reviewed for fidelity measurement, fidelity analysis reported, or interrater reliability reported.

Data collection is being described with more detail, but there is room for improvement. Meeting the indicator with high quality for data collection would indicate that articles stated specific information about how many times data were collected, who collected the information, and how the data were collected. As Gersten et al. (2005) pointed out, "whether" fidelity was measured is important, but so is the "how" it was measured. Fidelity data for early reading interventions is most often collected by means of direct observations, with 91% of the studies reporting fidelity claiming to use this method.

Fidelity monitoring is also happening at an increasing rate. An average score of 1.00 indicates that in most studies, the researchers monitored or collected data once or

twice throughout the intervention. Meeting this indicator with high quality (2.00) would have entailed monitoring fidelity multiple times. Gersten et al. (2005) recommended that observations take place on a regular basis over the entire intervention.

Descriptions of how fidelity is measured are important so that information can be gleaned regarding the ways in which fidelity was assessed. In the studies reviewed here, the measurement instruments used most often were rating scales and checklists. The rating scales ranged from 3-point to 5-point. For example, Kuhn et al. (2006) used a 5-point scale (1 = no fidelity to 5 = very high fidelity). Checklists dichotomously measured whether an aspect of the intervention was performed or not. The number of checklist items ranged from 8 (Lo et al., 2009) to 79 (Fuchs et al., 2008). Generally, a percentage of components performed was reported for the fidelity score.

Nine studies reported use of published observation protocols. Quality was evaluated using the Classroom Atmosphere Rating Scale (McMaster et al., 2008). Adherence to a fluency intervention was captured using the fluency-oriented reading instruction (Kuhn et al., 2006; Schwanenflugel et al., 2009) protocol. Program differentiation was evaluated using the Instructional Content Emphasis-Revised (Edmunds & Briggs, 2003) to capture the extent to which content and activities occurred in both experimental and control conditions (Denton et al., 2010; Vadasy & Sanders, 2008b; Vadasy & Sanders, 2010; Vadasy & Sanders, 2011).

Schulte et al. (2009) argued that "the development of a reliable, validated, generic treatment integrity instrument is another area in which education could build" (p. 469). For early reading interventions, a generic instrument could be created to capture

components that are often found in these interventions along with practices known to be evidence based. This type of instrument would be helpful in comparing studies (Schulte et al., 2009). A generic instrument would also prove to be a more cost-effective way to collect fidelity data because most instruments currently used are researcher developed, which takes time away from the actual intervention.

Among the early reading studies we reviewed, fidelity data analysis failed to yield a mean score of 1.00, indicating that this indicator was not adequately met. The score fell below a 1.00, signifying that the indicator had not been met, but they were close. Under the fidelity rubric, fidelity data analysis was scored a 1.00 if descriptives were reported and a 2.00 if fidelity was used as an independent variable. Even with only seven articles receiving a 2.00 for this category, the number is still promising. Articles are reporting descriptive data along with information about the fidelity measurement instrument much more frequently than expected.

Finally, the analysis of fidelity using the quality indicators revealed that the area in greatest need for improvement is interrater reliability reporting. As previously stated, direct observation was the most common data collection approach and rating scales were the most prominent measurement instrument. However, interrater reliability was not often reported. In fact, 70% of the articles failed to report any type of interrater reliability index. Those studies that did report interrater reliability calculated it using percent agreement, Cohen's Kappa, Pearson r correlations, and intraclass correlation coefficients (ICC). The time and personnel required to obtain interrater reliability on 20 to 25% of the data can be an expensive and inconvenient aspect of fidelity measurement.

Nonetheless, assessing interrater reliability is critical to ensure that fidelity assessments are free from bias and reflect reliable and valid information about an intervention.

Limitations

It is important to highlight the limitations of this study. First, although a comprehensive search was completed carefully over a 6-month period, there may be articles that were not identified. In addition, this review was restricted to early reading interventions that employed group experimental or quasi-experimental designs; thus, findings cannot be generalized to other fields of research or single-case studies.

While each study was carefully examined more than once to accurately discern fidelity measurement, it is possible that authors could have conducted or measured fidelity in ways that were not reported in the published articles. Because journals have page limits and fidelity can be a complicated subject, many articles may not have fully reported the entire fidelity story.

The articles were coded based upon predetermined criteria; as a result, information was gleaned from the articles as is and was not based upon any judgment of whether or not a particular fidelity dimension should have been recorded. For example, several of the studies employed computer-based interventions in which the software went through explicit steps with little room for deviation. In such cases, fidelity as an adherence dimension might not be applicable because the computer automatically and consistently performed with high fidelity.

Conclusion

In summary, fidelity of implementation is essential for making causal inferences about an intervention's effects or lack thereof. Understanding the entire story of an intervention plays an important role is determining quality research. The numerous calls for increased emphasis on fidelity of implementation this decade have increased the extent to which fidelity is being addressed and reported in early reading intervention articles. However, this study's findings suggest there is room for improvement regarding many of the fidelity quality indicators. Further research is needed to determine the best, most cost-effective approaches for measuring, analyzing, and reporting fidelity of implementation.

CHAPTER III

AN EXPLORATORY STUDY INVESTIGATING PROGRAM DIFFERENTIATION FOR AN EARLY READING INTERVENTION

Fidelity of implementation, synonymous with treatment fidelity or treatment integrity, refers to "the degree to which a treatment condition is implemented as intended" (Moncher & Prinz, 1991, p. 247). Measuring whether an intervention is implemented as intended is crucial to understanding and interpreting outcomes. According to Sheridan et al. (2009), "Our ability to infer that an intervention is effective requires knowledge about its implementation" (p. 477). Without assurance that the intervention was implemented with fidelity, the internal validity (or the ability to make casual statements about the intervention) is undermined (Gresham et al., 2000). Only after careful examination of what actually occurred during the intervention can researchers gain insight into "which elements led to student outcomes" (Gersten, Baker, & Lloyd, 2000, p. 4).

Recently there has been a call for educational researchers to measure and report fidelity of implementation more thoroughly (Sanetti & Kratochwill, 2009). In a special issue of *Exceptional Children*, Gersten et al. (2005) described quality indicators for group experimental and quasi-experimental research and established essential and desirable features necessary to deem research of high enough quality to be considered evidence-based. The fidelity indicator provided by Gersten et al. advocated that a specific question—"was the fidelity of implementation described and assessed?" (p. 152)—be addressed as an essential indicator of quality group experimental research.

Furthermore, Gersten et al. suggested that researchers assess fidelity multiple times throughout the intervention, at a minimum use a checklist to capture the key components of the intervention, include some type of interobserver score, and ensure adequate time and intervention coverage was provided to the learner.

Many dimensions have been proposed to capture treatment fidelity, such as adherence/surface, dosage, quality/process, program differentiation, and participant responsiveness (Dane & Schneider, 1988; Gersten et al., 2005). The adherence dimension assesses whether specified components of the intervention were delivered as prescribed (Dane & Schneider, 1988; Gersten et al., 2005). In a comprehensive search of group experimental kindergarten through third grade early reading interventions from the years 2005-2011, 84 articles were selected and coded for different dimensions of fidelity (Fogarty, Simmons, & Hagan-Burke et al., 2012). The most common fidelity dimension reported was adherence, with 58 (69%) of the 84 articles reporting adherence.

Of the 84 articles selected in the comprehensive search of K-3 intervention research, only 10 studies (11.9%) included fidelity measures to capture the difference between treatment and comparison groups (Fogarty et al., 2012). Therefore, the fidelity focus of the current study is program differentiation, in an effort to add to the extant literature by exploring methods of measuring and computing program differentiation.

Program differentiation, sometimes called treatment differentiation, delineates "whether treatment conditions differ from one another in the intended manner" (Moncher & Prinz, 1991, p. 248). Program differentiation answers the following

question: how alike and different are the conditions in the study? Program differentiation permits researchers to determine the differences between the comparison treatments, in order to confidently demonstrate the impact of the intervention (Century et al., 2010). Another important reason to assess program differentiation is to assure that intervention drift is kept at a minimum, especially if an interventionist was assigned to both conditions for a within-teacher design (Dane & Schneider, 1998).

Program differentiation is particularly relevant to early reading intervention and the converging evidence base that has accumulated over the past two decades. The body of research in early reading intervention suggests that effective interventions emphasize essential components of reading such as vocabulary, decoding, phonemic awareness, comprehension, and fluency (Gersten et al., 2005; National Reading Panel, 2000; Scamacca et al., 2007). Evidence-based practices that have been identified by past research further indicate the significance of intervention delivery features. Delivery features such as explicit and direct instruction, intensive instruction through small groups and/or more instructional time, plentiful opportunities to respond, appropriate pacing, and supportive instruction through scaffolding and feedback, have been documented as essential intervention features (Bursuck & Blanks, 2010; Carnine, Silbert, Kame'enui, Tarver, & Jungjohann, 2006; Foorman et al., 2003; Foorman & Torgesen, 2001; Scammaca et al., 2007).

While essential features of effective early intervention have been identified, few studies have compared different interventions according to essential features. The concept of program differentiation, or the comparison of different interventions

according to a common set of features, has been proposed as a method to study fidelity and understand intervention effects (Hulleman & Cordray, 2009). The current study was designed to examine program differentiation between two interventions using a common fidelity measure based on research-based components of effective early reading interventions. The fidelity measure was used to compare and contrast the sameness and differences between the experimental condition and the comparison early reading intervention.

As context for this research, the following is a review of early reading intervention research examining program differentiation.

Program Differentiation in Early Reading Intervention Research

Even though Gersten et al. (2005) described the process of assessing and describing comparison condition data as the "least glamorous and most neglected" (p. 158) aspect of research, they advocated that the following question be addressed: "Was the nature of services provided in comparison condition described?" (p. 152). Reporting the procedures of the comparison condition is fundamental to program differentiation. The following examples illustrate procedures used by researchers to compare essential features of early reading interventions.

In a Tier 2 vocabulary and decoding intervention for kindergarten, Nelson et al. (2011) observed tutors in both treatment and control conditions, scoring on a 5-point rating scale general teaching behaviors. The general teaching behaviors consisted of (a) "whether book or props were visible to all children," (b) "whether all children were responsive during the session," and (c) "whether the children were appropriately kept on

task" (p. 195). A comparison of the means of the treatment and control groups revealed no statistically significant differences between the two groups.

In a first-grade supplemental reading intervention, Denton et al. (2010) observed the intervention and typical practice using the Instructional Content Emphasis-Revised (Edmunds & Briggs, 2003) observation protocol to collect data on the type of content taught in both conditions. Descriptive data were reported on content areas such as comprehension and phonological awareness. An example of the information reported was that the treatment condition spent 30.4% of time allocated to fluency, while the typical school practice had 13.5% of time allocated to fluency. Similarly, Apthorp (2006) developed a classroom observation protocol to document the nature and content of instruction in both the treatment and comparison condition of a supplemental vocabulary intervention third grade. Apthorp used the resulting information to confirm that teachers in the treatment condition taught vocabulary more frequently than those in the control condition.

Hulleman and Cordray (as cited in Cordray and Pion, 2006) referred to program differentiation as treatment differentiation and noted that the treatment "has to be stronger than or different from the counterfactual condition" (p. 91) on a parallel assessment. Using different indices or dimensions such as dose and participant responsiveness, Hulleman and Cordray (2009) used data collected from both conditions to create an achieved relative strength (ARS) index. The ARS index is the difference between what was implemented in the treatment group versus what was implemented in the control group and can be compared to conventional effect sizes such as Hedges's g.

Hulleman and Cordray (2009) presented an example of how the achieved relative strength index can be calculated and analyzed using the student responsiveness dimension. The study examined a motivational intervention by asking students to respond in writing to a prompt. Students in the treatment condition were asked to write how a math activity was relevant to their lives. Students in the control condition were asked to simply summarize the topic. The study was first conducted in a laboratory with highly controlled conditions and later transferred to high school classrooms. Achieved relative strength indexes were computed to compare the differences between the lab and the classroom condition on a 4-point scale designed to gather quantity and quality of student response. Using an average fidelity score for the laboratory and classroom conditions, an effect size of 1.20 was calculated, revealing that the magnitude or strength of participant responsiveness was quite large for the laboratory condition.

Present Study

Using exploratory factor analysis, this study sought to examine whether and to what extent two early reading interventions differed on evidence-based features of effective instruction. To study program differentiation, the study used the Fidelity Observation Guide to document instructional practices in both conditions. Using data collected from the FOG at three points during the intervention period, the researcher identified latent factors of effective reading practices and tested whether experimental and typical practice conditions differed on these constructs. This study sought to answer the following research questions:

1. Do indicators intended to measure effective teaching practices create latent factors as measured by exploratory factor analysis?
2. Using latent variables of instruction, to what extent do interventions differ as indexed by an achieved relative strength index?

Method

Data for this study came from a larger program of intervention research that examined variations of early reading interventions and their effects on the reading achievement of kindergartners identified as at risk for reading difficulties (Simmons et al., 2011). Data were from one of the experimental study years that compared the effects of researcher-developed and school-designed RTI approaches to early reading intervention.

Researcher-Developed Early Reading Intervention

The researcher-developed response to intervention approach used the Early Reading Intervention (ERI; Pearson/Scott Foresman, 2004) as the base curriculum. ERI has four primary parts that reflect evidence-based skills and methods to promote early reading success. It is intended to be implemented in small groups for 30 minutes a day, and in an RTI framework as a Tier 2 intervention. The intervention is a scripted program that explicitly and systematically teaches phonemic awareness, alphabetic skills, decoding, word identification, and sentence reading to kindergartners. Each lesson contains seven to eight activities that last from 1 minute to 8 minutes. Activities 1 and 2 focus on phonemic and alphabetic skills. Activities 3, 4, and 6 integrate phonologic

awareness skills with alphabetic components. Activities 5, 7, and 8 integrate writing and spelling. Skills in the program increase in complexity across the 126 lessons.

The ERI curriculum uses the model, lead, test method of teaching and specifies the language to introduce skills. For example, when introducing an initial sound game, the interventionist is prompted to say, "My turn. I'll say the name of the picture and then tell whether it begins with /k/: cat. Cat begins with /k/" (Pearson/Scott Foresman, 2004, p. 75). The ERI program includes directions for feedback by reminding the interventionist to "confirm correct response and prompt sound production" (Pearson/Scott Foresman, 2004, p. 177). Modification procedures are also embedded for students who need them.

School-Designed Early Reading Intervention

The comparison intervention consisted of school-designed instruction (SDI) implemented by reading interventionists. The interventionists were allowed to determine the content of instruction as well as delivery features. Most interventionists in the SDI group used a guided reading approach with connected text to read with students. Two of the six SDI interventionists implemented commercial intervention programs—*Reading Mastery Plus* (Engelmann & Hanner, 2001) and *Road to the Code* (Blachman, Ball, Black, & Tangel, 2000). The other four SDI interventionists reported that the implemented interventions were teacher developed and included the district's core curriculum materials.

The SDI interventionists received no additional professional development. However, all interventionists in participating schools were experienced in providing

supplemental beginning reading intervention and had previously received professional development and resources in evidence-based reading instruction methods.

Systematic RTI methods were in place in all participating schools at varying levels of implementation. These procedures focused on formative assessment, grade-level data meetings, and Tier 2 intervention support. In order to be able to compare the two groups, the ERI and SDI conditions were similar in a number of ways: (a) interventions were held in small groups, (b) groups consisted of five students, (c) groups met every day for 30 minutes, (d) each intervention was implemented for the same number of lessons, and (e) each focused on early literacy skills.

Participants

Interventionists. Reading interventionists were assigned within their school to either the ERI (n = 15) or SDI conditions (n = 5), as displayed in Table 3.1. One interventionist participated in both conditions. ERI interventionists received two formal professional development sessions. One full-day session occurred in the fall before implementation, covering the first and second part of the ERI program. Another full-day session occurred in late January for the remainder of the intervention. Fidelity checks were conducted three times throughout the year and used to provide feedback to the interventionist.

Table 3.1

Interventionist Demographics

Variable	n	%
Female Teachers	20	100
Highest Degree Earned:		
High School/Less than Bachelor's	3	10
Bachelor's	9	30
Master's	15	50
Ed.S./Doctorate	2	6.6
Total Years of Teaching Experience:		
Mean	12.77	
SD	7.68	
Years of Experience Teaching Kindergarten:		
Mean	12.10	
SD	8.10	

Students. Children selected for the study were screened at the beginning of kindergarten. In the first several weeks of the school year, teachers selected five to eight students based on previously collected school data (e.g., DIBELS). Research staff conducted further assessment on the remaining students. To qualify, students had to perform ≤ the 36th percentile on DIBELS letter naming fluency, ≤ 37th percentile on CTOPP sound matching, ≤ the 9th percentile on the letter/word identification subtest of the WRMT-R/NU (1987, 1998), and/or ≤ the 16th percentile on the CTOPP rapid object naming subtest. The students then received the ERI intervention every day for 30 minutes. Table 3.2 describes the student demographics.

Table 3.2

Student Demographics

Variable	(n = 67) Treatment		(n = 23) Control	
	n	%	n	%
Gender:				
Male	37	55.2	14	60.9
Female	30	44.8	9	39.1
Ethnicity:				
Black or African-American	25	37.3	5	21.7
Hispanic or Latino	13	19.4	8	34.8
White	24	35.8	10	43.5
Other	5	7.5	0	0
Recipient of Special Education Services	6	9.0	3	13.0
Bilingual/English Language Learner	11	16.4	5	21.7
Age:				
Mean	5.55		5.5	
SD	.28		.34	

Measures

The Fidelity Observation Guide was created expressly for this study (see Appendix C). The Fidelity Observation Guide began as a hybrid of the Reading in Special Education (RISE; Brownell et al., 2009) and the English Language Learner Classroom Observation instruments (Baker, Gersten, Haager, & Dingle, 2006). The indicators from these observation protocols were chosen based on their compatibility with the ERI program.

The original indicators selected from the RISE included (a) “extent to which students are highly engaged,” (b) “provides explicit instruction,” (c) “effectively

redirects and proactively address behavior," and (d) "creates warm and supportive environment" (Brownell, 2009, p. 401). The original indicators from the English-Language Learner Classroom Observation Instrument included (a) "models new skills and strategies during lesson," (b) "gives feedback on academic performance," (c) "modifies instruction for students as needed," (d) "uses visuals or manipulatives to teach content," and (e) "provides systematic instruction" (Baker et al., 2006, p. 117).

After several iterations, the following three domains and indicators were identified and finalized for the Fidelity Observation Guide: (a) instructional practices, (b) teacher responsiveness, and (c) student engagement. These domains constituted indicators that are considered essential features of the ERI curriculum. Because ERI is a manualized intervention with explicit components and scripted lessons, the indicators of the FOG were expected to be implemented at high levels (Lane, Bocian, MacMillan, & Gresham, 2004).

The instructional practice domain consisted of four items: (a) models new skills and strategies during lesson, (b) provides guided practice, (c) uses manipulatives/hands-on opportunities during instruction, and (d) provides a systematic instruction that follows an instructional sequence to the intervention. The teacher responsiveness domain included (a) provides appropriate/fluent pacing, (b) provides immediate feedback to students, (c) modifies and/or scaffolds instruction as necessary, and (d) provides plentiful opportunities for students to respond. The student engagement domain included (a) students are engaged, and (b) students refrain from problem behavior. Appendix D

provides the codebook for the FOG and contains examples and non-examples of each indicator used during fidelity training.

Although a low inference measurement, such as a checklist, is ideal for reliability and ease of use, this study wanted to be able to capture as much variability as possible. Therefore, the FOG was scored using a 4-point Likert scale (0 = not observed, 1 = observed with low quality/inconsistent high, 2 = observed with adequate quality, 3 = observed consistently with exceptional quality).

Data Collection

Three times throughout the school year, interventionists from both the treatment and comparison conditions were observed and videotaped. The first author and a project director later coded the video tapes using the FOG for each observation period. The interventionists were randomly assigned a coder.

Interrater reliability was established prior to the study through video tapes of interventionists teaching the ERI curriculum from two separate studies. The interrater reliability rate was established at 80%. Ongoing interrater reliability checks occurred throughout the data collection period. Twenty-two percent of instructional sessions were coded, and the overall interrater reliability percent agreement was 92.6%. Interrater reliability for each domain was (a) 92.4% for instructional practices, (b) 92.7% for teacher responsiveness, and (c) 94.6% for student engagement.

Data Analysis

The main purpose of the analysis was to evaluate how well the indictors of the FOG reflected multiple dimensions of fidelity. Exploratory factor analysis (EFA) was

used to identify latent instructional constructs at each observation point. Findings are reported by time period. EFA was also conducted to determine whether the indicators that were expected to converge actually did. EFA is an analysis that measures the relationship between observable indicators (such as the items of the FOG) to determine a smaller number of latent constructs or factors (Thompson, 2004). When conducting an EFA, predetermined factors are not required. The crucial step in conducting an EFA is to examine the relation of items to latent factors and then identify the latent factors based on a theory.

Principal component analysis was used to extract the factors by placing 1.0s on the diagonal of the correlation matrix (Thompson, 2004). To determine the number of factors on the FOG, the Kaiser-Guttman criteria of eigenvalues > 1.0 was employed. Pattern and structure coefficients were used to analyze the factor variables once the Promax rotation was conducted.

Fidelity at each observation time point was analyzed separately. The sample at each time point was as follows: (a) the first time point had 21 observations (n = 21); (b) the second time point had 22 observations (n = 22); and (c) the third time point had 23 observations (n = 23). After the initial EFA was conducted using the first time point, confirmatory factor analysis was conducted using the same factors in the second and third time points.

Effect Sizes

Achieved relative strength index was calculated using Cordray's formula:

$$ARS\ Index = \frac{\mathbf{t}^{Tx} - \mathbf{t}^{C}}{S}$$

This formula is an effect size using Hedge's g formula (Hedges, 2007). The numerator is the difference between the treatment and control on some aspect of the intervention. For the purpose of this study, latent variables were created based on the results of the exploratory factor. The denominator is the pooled standard deviation from the treatment and control group. Using the average fidelity index approach (Hulleman & Cordray, 2009), each latent construct was averaged for each condition. Next, an effect size was created at each observation time point. The effect sizes created using this formula were the magnitude of the differences between experimental and control conditions.

Results

Exploratory Factor Analysis

Observation 1 results. Observation 1 produced three factors with eigenvalues over 1.0, establishing a three-factor model. The first factor, instructional practices, had an eigenvalue of 4.73 and accounted for 47.0% variance. For the second factor, teacher responsiveness, and the third factor, student engagement, eigenvalues were smaller, at 2.00 and 1.15, and accounted for 20.0% and 11.5% variance, respectively.

As seen in Table 3.3, indicators and their pattern coefficients for instructional practices were (a) provides guided practice (.775), (b) models new skills (.823), (c) provides systematic instruction (.836), and (d) uses manipulatives (.959). Indicators and their corresponding path coefficients for teacher responsiveness included (a) provides immediate feedback (.663), (b) uses appropriate pacing (.688), (c) provides opportunity to respond (.700), and (d) modifies and/or scaffolds instruction as necessary (.952).

Finally, student engagement had two indicators: (a) students refrain from problem behavior (.827), and (b) students are engaged (.921). Table 3.4 displays the factor correlations of .196, .356, and .383.

Table 3.3

Rotated Principal Component Analysis Pattern/Structure Coefficients for Observation 1

	Pattern			Structure			
Variable	I	II	III	I	II	III	h^2
Uses manipulatives	**.959**	-.384	.244	.859	.070	.295	.869
Systematic instruction	**.836**	-.097	.118	.822	.265	.247	.690
Models new skills	**.823**	.304	-.121	.915	.576	.148	.910
Guided practice	**.775**	.345	-.241	.860	.556	.033	.851
Modifies instruction	-.005	**.952**	-.094	.341	.917	.244	.848
Opp. to respond	.109	**.700**	.121	.401	.785	.391	.640
Appropriate pacing	-.293	**.688**	.264	.022	.670	.452	.574
Immediate feedback	.125	**.663**	.269	.241	.451	.966	.731
Students engaged	.016	.117	**.921**	.241	.451	.966	.946
Students no problems	.081	.133	**.827**	.294	.459	.890	.821

Note. Pattern coefficients greater than |.45| are in boldface.

Table 3.4

Factor Correlation Matrix for Observation 1

Component	I	II	III
1	1.000	.383	.196
2	.383	1.000	.356
3	.196	.356	1.000

Observation 2 results. Observation 2 also produced three factors with eigenvalues over 1.0, verifying a three-factor model. The first factor, instructional practices, had an eigenvalue of 3.96 and accounted for 39.6% variance. Eigenvalues for the second factor, student engagement, and the third factor, teacher responsiveness, were smaller, at 2.18 and 1.33, and accounted for 21.8% and 13.3% variance, respectively.

As seen in Table 3.5, indicators and their pattern coefficients on instructional practices were (a) provides guided practice (.813), (b) provides systematic instruction (.827), (c) models new skills (.889), and (d) uses manipulatives (.922). The second factor, student engagement, had three indicators: (a) uses appropriate pacing (.762), (b) students are engaged (.973), and (c) students refrain from problem behavior (.990). Indicators and their corresponding path coefficients for teacher responsiveness included (a) provides immediate feedback (.557), (b) provides opportunity to respond (.615), and (c) modifies and/or scaffolds instruction as necessary (.890). Table 3.6 displays the factor correlation scores, which were .260, .276, and .277.

For Observation 2, most of the indicators were associated with the same factors as in Observation 1. However, there was one notable change. The indicator “uses appropriate pacing” moved from the teacher responsiveness factor to the student engagement factor.

Table 3.5

Rotated Principal Component Analysis Pattern/Structure Coefficients for Observation 2

	Pattern			Structure			
Variable	I	II	III	I	II	III	h^2
Uses manipulatives	**.922**	-.187	-.196	.819	-.002	.007	.755
Systematic instruction	**.827**	.091	.036	.860	.316	.290	.750
Models new skills	**.889**	.023	-.078	.873	.232	.174	.768
Guided practice	**.813**	-.025	.112	.837	.217	.329	.711
Modifies instruction	-.327	-.124	**.890**	-.113	.038	.766	.848
Opp. to respond	.373	.118	**.615**	.573	.386	.751	.722
Appropriate pacing	-.064	**.762**	.198	.189	.800	.391	.675
Immediate feedback	.341	.036	**.557**	.504	.279	.661	.550
Students engaged	-.074	**.973**	-.013	.176	.950	.237	.908
Students no problems	.034	**.990**	-.215	.232	.939	.069	.923

Note. Pattern coefficients greater than |.45| are in boldface. Sorted based on pattern coefficients from Observation 1.

Table 3.6

Factor Correlation Matrix for Observation 2

Component	I	II	III
1	1.000	.260	.276
2	.260	1.000	.277
3	.276	.277	1.000

Observation 3 results. Observation 3 also produced three factors with eigenvalues over 1.0, again verifying a three-factor model. The first factor, instructional practices, had an eigenvalue of 3.20 and accounted for 32.0% variance. Eigenvalues of the second factor, student engagement, and the third factor, teacher responsiveness, were smaller, at 2.68 and 1.27, and accounted for 26.6% and 12.7% variance, respectively.

As seen in Table 3.7, indicators and their path coefficients on instructional practices were (a) provides guided practice (.660), (b) models new skills (.702), (c) uses manipulatives (.873), and (d) provides systematic instruction (.927). Student engagement had three indicators: (a) uses appropriate pacing (.527), (b) students are engaged (.965), and (c) students refrain from problem behavior (.928). Indicators and their corresponding path coefficients for teacher responsiveness included (a) provides immediate feedback (.557), (b) modifies and/or scaffolds instruction as necessary (.595), and (c) provides opportunity to respond (.869). Table 3.8 displays the factor correlations of -.071, .176, and .319.

Observation 3 aligned more with the results of Observation 2 in that the indicator "uses appropriate pacing" again moved from the teacher responsiveness factor to the student engagement factor. The indicator "models new skills" had a path coefficient for of .702 for instructional practices and .463 for teacher responsiveness, indicating that it was correlated with both factors.

Table 3.7

Rotated Principal Component Analysis Pattern/Structure Coefficients for Observation 3

	Pattern			Structure			
Variable	I	II	III	I	II	III	h^2
Uses manipulatives	**.873**	.147	-.274	.815	-.003	-.073	.731
Systematic instruction	**.927**	.087	.018	.924	.026	.209	.863
Models new skills	**.702**	-.047	**.463**	.590	.143	.768	.806
Guided practice	**.660**	-.198	.391	.742	-.120	.444	.687
Modifies instruction	**-.564**	.092	**.595**	-.466	.322	.526	.605
Opp. to respond	-.184	.006	**.869**	-.032	.296	.838	.736
Appropriate pacing	.247	**.527**	.039	.217	.521	.251	.338
Immediate feedback	.191	.334	**.504**	.256	.481	.644	.534
Students engaged	-.017	**.965**	.009	-.085	.969	.313	.940
Students no problems	-.070	**.928**	**.055**	-.127	.951	.339	.910

Note. Pattern coefficients greater than |.45| are in boldface. Sorted based on pattern coefficients from Observation 1.

Table 3.8

Factor Correlation Matrix for Observation 3

Component	I	II	III
1	1.000	-.071	.176
2	-.071	1.000	.319
3	.176	.319	1.000

Confirmatory Factor Analysis

Attempts were made to run a confirmatory factor analysis for Observations 2 and 3 with the model derived from the exploratory factor analysis of Observation 1. Due to the small sample size, the model indices never fit appropriately. Appendix E displays the results and model indices.

Effect Sizes

Achieved relative strength indexes were computed for each latent construct (see Table 3.9). For Observation 1, the ARS indexes were high and ranged from .23 to 2.20, indicating that ERI teachers on average differed from the control conditions on the latent constructs formed from the Fidelity Observation Guide. For Observation 2, the scores ranged from -.04 to 1.92. The -.04 effect size was for the student engagement construct and indicated that the control condition students were on average more responsive at the second time point. The third observation effect sizes ranged from -.27 to 1.64. This time teacher responsiveness had an effect size of -.52 and student engagement had an effect size of -.27, which indicated that the control interventions on average had stronger scores in these domains than the treatment interventionist.

Table 3.9

Effect Sizes on Latent Constructs (The Difference Between Treatment and Control on Each Latent Construct-Program Differentiation)

	Observation		
Variable	I	II	III
Instructional Practices	2.20	1.92	1.64
Teacher Responsiveness	.41	.06	-.52
Student Engagement	.23	-.04	-.27

Discussion

For the purpose of this study, the FOG was used to establish program differentiation between an experimental condition using the ERI curriculum and a school-designed intervention. Program differentiation is a necessary dimension of fidelity for establishing a more confident causal statement about the treatment effects. This study sought to extend the research regarding the fidelity dimension of program differentiation by developing a fidelity instrument that could be used to evaluate common components between treatment and comparison conditions.

The Fidelity Observation Guide was based on previously validated instruments (Brownell et al., 2009; Haager et al., 2001) and was designed to capture essential features of effective instruction. For the first observation period, the variables that made up the instructional practice domain included (a) provides guided practice, (b) models new skills, (c) provides systematic instruction, and (d) uses manipulatives. The teacher responsiveness factor included (a) provides immediate feedback, (b) uses appropriate pacing, (c) provides opportunity to respond, and (d) modifies instruction as necessary. The student engagement factor included (a) students are engaged in task, and (b) students refrain from problem behavior.

For Observations 1 and 3, the factors changed slightly. The first factor, instructional practice, remained the same as in Observation 1. The teacher responsiveness factor for both Observations 1 and 2 included (a) modifies instruction when necessary, (b) provides opportunities to respond, and (c) provides immediate feedback. The student engagement factor changed slightly for Observations 2 and 3. The

factors included (a) students are engaged, (b) students refrain from problem behavior, and (c) appropriate pacing is used.

Although the pacing indicator does not seem to fit in the student engagement factor, it makes sense that pacing is strongly correlated with engagement. The better the pacing, the more likely the students are to remain engaged and out of trouble. If the teacher is teaching too slowly, the students will likely become bored. If the teacher is moving too quickly, the students will become frustrated. Pacing is a Goldilocks skill that requires "just right" timing.

After the dimensions were created, for each time point, the achieved relative strength index was created for each latent construct: (a) instructional practices, (b) teacher responsiveness, and (c) student engagement. The ARS is modeled after a Hedge's g effect size and for the purposes of this study was the magnitude of difference between the causal components of both conditions.

Overall, at all three time points, the instructional practice domain had a large effect size (2.20, 1.64, and 1.92). These effects sizes indicate that the treatment groups were very different from the control group on the latent construct instructional practices. This finding is significant because it demonstrates that the level of instructional practices (e.g., modeling, systematic instruction, guided practice, and manipulatives) in the treatment group was stronger than in the comparison group. Theoretically, the strength of intervention and program differentiation is strong and should impact student outcomes.

The teacher responsiveness effect sizes were .41, .06, and -.52. For Observation 1, the treatment interventionist performed better; however, the control interventionist performed better on this dimension during Observation 3. For Observation 2, the small effect size (.06) revealed that the two conditions were similar. Again, this finding is significant because the indicators in teacher responsiveness are considered essential items of evidence-based practices and modeled in ERI. These instructional behaviors are also difficult to train on because the teacher has to perform these as needed and in direct response to what the student does. The findings suggest that the performance on this construct was about equal and should not impact student outcomes.

Finally, the student engagement effect sizes of .23, -.04, and -.27 reveal that the magnitude of difference between the two conditions is small. Again, this indicates that program differentiation for the student engagement construct is insignificant and should not impact student outcomes.

Limitations

Sample size severely limited the analyses for this study. Therefore, the main limitation of this study is the relatively small number of observations relative to the analyses performed. Perhaps that is why fidelity measures are so difficult to create and to examine. Collecting data for fidelity assessment is expensive and time consuming. However, to be able to understand fidelity and the multiple dimensions proposed, larger numbers of observations may be needed.

Conclusion

In an effort to add to the literature on fidelity of implementation, a Fidelity Observation Guide was created to capture evidence-based practices during early reading interventions, with the intent to compare two comparison conditions. Program differentiation is an important aspect of calculating fidelity of implementation for a clearer picture of the differences that can affect student outcomes.

CHAPTER IV

CONCLUSION

Fidelity of implementation is receiving increased attention as educational researchers strive to identify evidence-based practices and discern the conditions under which those practices are most effective. Measures of implementation fidelity can increase one's confidence that results of a study are directly related to the corresponding intervention. Well-documented fidelity may also provide insight regarding specific elements of an intervention that were most strongly associated with student outcomes (Gersten et al., 2000). Without assurance that an intervention was implemented with fidelity, the internal validity (i.e., ability to make casual statements about the intervention) is undermined (Gresham et al., 2000). Fidelity is also crucial for understanding the absence of intervention effects.

The purpose of this dissertation was to examine fidelity of early reading interventions through (a) reviewing fidelity reporting in early reading research published between 2005-2011, and (b) examining program differentiation by developing a fidelity instrument intended to measure essential early reading practices. The first study sought to determine the state of fidelity reporting, and the second study sought to examine whether and to what extent two early reading interventions differed on evidence-based features of effective instruction. Following is a summary of the conclusions as well as implications for future research.

Study 1: Examining Fidelity of Implementation in Current Early Reading Intervention Research

Concerning the prevalence and types of fidelity reported in early reading intervention research, findings indicate that fidelity reporting has increased over the past 6 years; however, reporting is not equal for all dimensions. Different dimensions of fidelity have been proposed, such as adherence/surface, dosage, quality/process, program differentiation, and participant responsiveness (Dane & Schneider, 1998; Gersten et al., 2005). Of the dimensions, adherence was the most widely reported dimension, with 58 of the 84 early reading research articles reporting some type of adherence score. Program differentiation was the least reported dimension. Based on findings, an important implication is the need for further examination of fidelity more comprehensively, including the different dimensions.

Extending and examining the relation between fidelity and outcomes is an area that needs to be further researched. To date, very few studies have directly examined the relationship between fidelity and student outcomes (O'Donnell, 2008). Of the 84 articles coded for this dissertation, only seven, or 8.6%, reported fidelity linked with student outcomes. Adherence and its relation to student outcomes produced mixed results. Most studies found little or no relation between adherence and student outcomes (Ehri et al., 2007; Gunn et al., 2011; Nelson et al., 2011; Simmons et al., 2007; Vadasy et al., 2006a). However, two studies found that adherence predicted student outcomes (Frechtling et al., 2006; Vadasy & Sanders, 2011).

While there is general agreement regarding its importance, there is a paucity of effective and efficient methods for measuring and reporting fidelity, particularly as it relates to student outcomes (O'Donnell, 2008). As greater emphasis is placed on treatment integrity in educational research, it is important to understand the components of an intervention that matter, as well as their relative contributions to student achievement.

In addition, findings from the review of research indicate a need for future research examining the frequency and methods used to study fidelity. The review revealed considerable variability in the frequency of fidelity measurement. Some studies specified the number of times that fidelity was monitored—reporting ranges from one to 26 times—while other studies reported percentages of times monitored—ranging from 10% to 56%. In addition, this study found that researchers use several different methods, such as checklists, rating scales, and published observation protocols. The field is not consistent in the frequency or methods used to measure fidelity (Power, Blom-Hoffman, Clark, Rilley-Tillman, Kelleher, & Manz, 2005). Further research is needed to find a parsimonious method of fidelity data collection.

Study 2: An Exploratory Study Investigating Program Differentiation for an Early Reading Intervention

To better understand intervention effects, it is important to disaggregate the features of intervention that are responsible for effects. While program differentiation provides valuable information to advance our understanding of intervention effects, it is rarely reported in early intervention research. The purpose of the second study was to

examine program differentiation between two interventions using a common fidelity measure based on research-based components of effective early reading interventions.

This study attempted to extend the research on the fidelity dimension of program differentiation by developing a fidelity instrument that could be used to assess commonalities between treatment and comparison conditions. The Fidelity Observation Guide was created to capture essential features of effective instruction. Using exploratory factor analysis, three latent constructs were created: (a) instructional practice, (b) teacher responsiveness, and (c) student responsiveness. Due to the small sample size, a confirmatory factor analysis did not produce a model with adequate fit; however, through exploratory factor analysis, these constructs remained fairly consistent.

To examine the strength of program differences, the achieved relative strength index was employed (Hulleman & Cordray, 2009). As indicated by findings, differences in instructional practices were more pronounced than differences in teacher responsiveness and student engagement. These findings, nonetheless, are insufficient to draw causal conclusions, as they were not linked to student outcomes.

Implications for Future Research

This dissertation was designed to add to the existent literature by examining the state of fidelity reporting in early reading interventions and investigating an application of program differentiation. While professional organizations and research standards advocate the need for more rigorous and complete fidelity reporting, findings here indicate considerable need for future research. In particular, further research is needed to

understand the components of an intervention that matter to student outcomes, as well as the relative contribution of each to student achievement. Fidelity of implementation procedures are resource intensive, and research that can help prioritize dimensions most associated with effects is particularly important. Moreover, it would be worthwhile to extend the investigation of program differentiation, as it may be particularly helpful in understanding intervention dimensions that are most responsible for effects, or the absence of effects. Finally, future research is needed on the methods of studying fidelity. Currently, there is considerable variability among processes used to measure fidelity, and examining questions of how often, for what duration, and on what scale to measure fidelity are particularly important if we are to systematically advance our understanding of fidelity of implementation.

REFERENCES

Amendum, S. J., Vernon-Feagans, L., & Ginsberg, M. C. (2011). The effectiveness of a technologically facilitated classroom-based early reading intervention. *The Elementary School Journal, 112*(1), 107-131.

Apthorp, H. S. (2006). Effects of a supplemental vocabulary program in third-grade reading/language arts. *Journal of Educational Research, 100*(2), 67-79. doi:10.3200/JOER.100.2.67-79

Baker, S. K., Gersten, R., Haager, D., & Dingle, M. (2006). Teaching practice and the reading growth of first-grade english learners: Validation of an observation instrument. *Elementary School Journal, 107*(2), 199-219.

Begeny, J. C., Laugle, K. M., Krouse, H. E., Lynn, A. E., Tayrose, M. P., & Stage, S. A. (2010). A control-group comparison of two reading fluency programs: The Helping Early Literacy with Practice Strategies (HELPS) program and the Great Leaps K-2 reading program. *School Psychology Review, 39*(1), 137-155.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246. doi:10.1037/0033-2909.107.2.238

Blachman, B. A., Ball, E. W., Black, R., & Tangel, D. (2000). *Road to the code: A phonological awareness program for young children.* Baltimore, MD: Brookes Publishing.

Brownell, M. T., Bishop, A. G., Gersten, R., Klingner, J. K., Penfield, R. D., Dimino, J., . . . Sindelar, P. T. (2009). The role of domain expertise beginning special education teacher quality. *Exceptional Children, 75*(4), 391-411.

Bursuck, B., & Blanks, B. (2010). Evidence-based early reading practices within a response to intervention system. *Psychology in the Schools, 47*(5), 421-431.

Calhoon, M. B., Al Otaiba, S., Greenberg, D., King, A., & Avalos, A. (2006). Improving reading skills in predominantly Hispanic Title 1 first-grade classrooms: The promise of peer-assisted learning strategies. *Learning Disabilities Research & Practice, 21*(4), 261-272. doi:10.1111/j.1540-5826.2006.00222.x

Carnine, D. W., Silbert, J., Kame'enui, E. J., Tarver, S. G., & Jungjohann, K. (2006). *Teaching struggling and at-risk readers: A direct instruction approach* . Upper Saddle River, NJ: Pearson.

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation, 31*, 199-218. doi:10.1177 /109821401036617

Chard, D. J., Ketterlin-Geller, L., Baker, S. K., Doabler, C., & Apichatabutra, C. (2009). Repeated reading interventions for students with learning disabilities: Status of the evidence. *Exceptional Children, 75*(3), 263-281.

Cordray, D., & Pion, G. (2006). *Treatment strength and integrity: Models and methods.* In R.R. Bootzin & P.E. McKnight (Eds.), *Strengthening research methodology: Psychological measurement and evaluation (pp.103-124).* Washington, DC: American Psychological Association. doi:10.1037/11384-006

Council for Exceptional Children (2008). *Classifying the state of evidence for special education practices: CEC practice study manual.* Arlington, VA, Shank, K.

Coyne, M. D., McCoach, D. B., & Kapp, S. (2007). Vocabulary intervention for kindergarten students: Comparing extended instruction to embedded instruction and incidental exposure. *Learning Disability Quarterly, 30*(2), 74-88. Retrieved from http://www.cldinternational.org

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli, R., Jr., & Kapp, S. (2009). Direct vocabulary instruction in kindergarten: Teaching for breadth versus depth. *Elementary School Journal, 110*(1), 1-18. doi:10.1086/598840

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*(1), 23-45. doi:10.1016/S0272-7358(97)00043-3

Denton, C. A., Kethley, C., Nimon, K., Kurz, T. B., Mathes, P. G., Minyi, S., & Swanson, E. A. (2010). Effectiveness of a supplemental early reading intervention scaled up in multiple schools. *Exceptional Children, 76*(4), 394-416. Retrieved from http://www.cec.sped.org

Edmunds, M., & Briggs, K. L. (2003). The instructional content emphasis instrument: Observations of reading instruction. In S. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: Systems for the observation of teaching and learning* (pp. 31-52). Baltimore, MD: Brookes.

Ehri, L. C., Dreyer, L. G., Flugman, B., & Gross, A. (2007). Reading rescue: An effective tutoring intervention model for language-minority students who are struggling readers in first grade. *American Educational Research Journal, 44*(2), 414-448. doi:10.3102/0002831207302175

Engelmann, S., & Hanner, S. (2001). *Reading Mastery Plus*. Columbus, OH: SRA/McGraw-Hill.

Fien, H., Santoro, L., Baker, S. K., Park, Y., Chard, D. J., Williams, S., & Haria, P. (2011). Enhancing teacher read alouds with small-group vocabulary instruction for students with low vocabulary in first-grade classrooms. *School Psychology Review, 40*(2), 307-318.

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., . . . Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science: The Official Journal of the Society for Prevention Research, 6*(3), 151-175.

Fogarty, M.S., Simmons, D.C, & Hagan-Burke, S. (2012). Examining fidelity of implementation in current early reading intervention research. Manuscript submitted for publication.

Foorman, B. R., Breier, J. I., & Fletcher, J. M. (2003). Interventions aimed at improving reading success: An evidence-based approach. *Developmental Neuropsychology, 24*(2-3), 613-639.

Foorman, B. R. T., & Torgesen, J. K. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research & Practice, 16*(4), 203.

Frechtling, J. A., Zhang, X., & Silverstein, G. (2006). The voyager universal literacy system: Results from a study of kindergarten students in inner-city schools.

Journal of Education for Students Placed at Risk (JESPAR), 11(1), 75-95. doi:10.1207/s15327671espr1101_5

Fuchs, D., Compton, D. L., Fuchs, L. S., Bryant, J., & Davis, G. N. (2008). Making "secondary intervention" work in a three-tier responsiveness-to-intervention model: Findings from the first-grade longitudinal reading study of the national research center on learning disabilities. *Reading and Writing: An Interdisciplinary Journal, 21*(4), 413-436. doi:10.1007/s11145-007-9083-9

Gersten, R., Baker, S., & Lloyd, J. W. (2000). Designing high-quality research in special education: Group experimental design. *Journal of Special Education, 34*(1), 2-12.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*(2), 149-164.

Good, R. H., & Kaminski, R. A. (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice, 15*(4), 198-205.

Gresham, F. M. (2009). Evolution of the treatment integrity concept: Current status and future directions. *School Psychology Review, 38*(4), 533-540.

Gunn, B., Smolkowski, K., & Vadasy, P. (2011). Evaluating the effectiveness of "read well kindergarten." *Journal of Research on Educational Effectiveness, 4*(1), 53-86. doi:10.1080/19345747.2010.488716

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics, 32*(4), 341-370.

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

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*(1), 88-110. doi:10.1080/19345740802539325

Institute of Education Sciences (IES). (2011). *Request for applications: Special education research grants*. Retrieved from http://ies.ed.gov/funding/pdf /2012_84324A.pdf

Jitendra, A. K., Burgess, C., & Gajria, M. (2011). Cognitive strategy instruction for improving expository text comprehension of students with learning disabilities: The quality of evidence. *Exceptional Children, 77*(2), 135-159.

Kamps, D., Abbot, M., Greenwood, C., Arreaga-Mayer, C., Wills, H., Longstaff, J., . . . Walton, C. (2007). Use of evidence-based, small-group reading instruction for English language learners in elementary grades: Secondary-tier intervention. *Learning Disability Quarterly, 30*(3), 153-168.

Kuhn, M. R., Schwanenflugel, P. J., Morris, R. D., Morrow, L. M., Woo, D. G., Meisinger, E. B., . . . Stahl, S. A. (2006). Teaching children to become fluent and automatic readers. *Journal of Literacy Research, 38*(4), 357-387. doi:10.1207/s15548430jlr3804_1

Lane, K. L., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2004). Treatment integrity: An essential—but often forgotten—component of school-based interventions. *Preventing School Failure, 48*(3), 36-43.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., . . . Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Clinical Trials, 6*(7), 1-28. doi:10.1371/journal.pmed.1000100

Lo, Y., Wang, C., & Haskell, S. (2009). Examining the impacts of early reading intervention on the growth rates in basic literacy skills of at-risk urban kindergarteners. *The Journal of Special Education, 43*(1), 12-28. doi:10.1177/0022466907313450

Mathes, P. G., Denton, C. A., Fletcher, J. M., Anthony, J. L., Francis, D. J., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly, 40*(2), 148-182.

Maynard, K. L., Pullen, P. C., & Coyne, M. D. (2010). Teaching vocabulary to first-grade students through repeated shared storybook reading: A comparison of rich

and basic instruction to incidental exposure. *Literacy Research and Instruction, 49*(3), 209-242. doi:10.1080/19388070902943245

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children, 71*(4), 445-463.

McMaster, K. L., Kung, S., Han, I., & Cao, M. (2008). Peer-assisted learning strategies: A "tier 1" approach to promoting English learners' response to intervention. *Exceptional Children, 74*(2), 194-214.

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*(3), 247-266. doi:10.1016/0272-7358(91)90103-2

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24*(3), 315-340.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

Nelson, J. R., Stage, S. A., Epstein, M. H., & Pierce, C. D. (2005). Effects of a prereading intervention on the literacy and social skills of children. *Exceptional Children, 72*(1), 29. Retrieved from http://www.cec.sped.org

Nelson, J. R., Vadasy, P. F., & Sanders, E. A. (2011). Efficacy of a Tier 2 supplemental root word vocabulary and decoding intervention with kindergarten Spanish-speaking English learners. *Journal of Literacy Research, 43*(2), 184-211.

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children, 71*(2), 137-148.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*(1), 33-84.

Pearson/Scott Foresman. (2004). *Scott Foresman sidewalks: Early reading intervention*. Glenview, IL: Author.

Power, T. J., Blom-Hoffman, J., Clarke, A. T., Riley-Tillman, T. C., Kelleher, C., & Manz, P. H. (2005). Reconceptualizing intervention integrity: A partnership-based framework for linking research with practice. *Psychology in the Schools, 42*(5), 495-507.

Puhalla, E. M. (2010). Enhancing the vocabulary knowledge of first-grade children with supplemental booster instruction. *Remedial and Special Education*. Advance online publication. doi:10.1177/0741932510362495

Rafdal, B. H., McMaster, K. L., McConnell, S. R., Fuchs, D., & Fuchs, L. S. (2011). The effectiveness of kindergarten peer-assisted learning strategies for students with disabilities. *Exceptional Children, 77*(3), 299-316.

Sanetti, L. M. H., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38*(4), 445-459.

Scammacca, N., Vaughn, S., Roberts, G., Wanzek, J., & Torgesen, J. K. (2007). *Extensive reading interventions in grades k– 3: From research to practice.* Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Schulte, A. C., Easton, J. E., & Parker, J. (2009). Advances in treatment integrity research: Multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treatment integrity. *School Psychology Review, 38*(4), 460-475.

Schwanenflugel, P. J., Kuhn, M. R., Morris, R. D., Morrow, L. M., Meisinger, E. B., Woo, D. G., . . . Sevcik, R. (2009). Insights into fluency instruction: Short- and long-term effects of two reading programs. *Literacy Research and Instruction, 48*(4), 318-336.

Sheridan, S. M., Swanger-Gagne, M., Welch, G. W., Kwon, K., & Garbacz, S. A. (2009). Fidelity measurement in consultation: Psychometric issues and preliminary examination. *School Psychology Review, 38*(4), 476-495.

Silverman, R. (2007). A comparison of three methods of vocabulary instruction during read-alouds in kindergarten. *Elementary School Journal, 108*(2), 97-113. doi:10.1086/525549

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

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

Stahl, S. A. & Heubach, K. (2005). Fluency-oriented reading instruction. *Journal of Literacy Research*, 37, 25-60.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. Multivariate Behavioral Research, 25(2), 173-180.

Taylor, B. M., & Pearson, P. D. (2000). *The CIERA School Change Classroom Observation Scheme.* Minneapolis: University of Minnesota.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

Vadasy, P. F., & Sanders, E. A. (2008a). Code-oriented instruction for kindergarten students at risk for reading difficulties: A replication and comparison of instructional groupings. *Reading and Writing: An Interdisciplinary Journal, 21*(9), 929-963. doi:10.1007/s11145-008-9119-9

Vadasy, P. F., & Sanders, E. A. (2008b). Repeated reading intervention: Outcomes and interactions with readers' skills and classroom instruction. *Journal of Educational Psychology, 100*(2), 272-290. doi:10.1037/0022-0663.100.2.272

Vadasy, P. F., & Sanders, E. A. (2009). Supplemental fluency intervention and determinants of reading outcomes. *Scientific Studies of Reading, 13*(5), 383-425.

Vadasy, P. F., & Sanders, E. A. (2010). Efficacy of supplemental phonics-based instruction for low-skilled kindergarteners in the context of language minority status and classroom phonics instruction. *Journal of Educational Psychology, 102*(4), 786-803. doi:10.1037/a0019639

Vadasy, P. F., & Sanders, E. A. (2011). Efficacy of supplemental phonics-based instruction for low-skilled first graders: How language minority status and pretest characteristics moderate treatment response. *Scientific Studies of Reading, 15*(6), 471-497. doi:10.1080/10888438.2010.501091

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006a). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. *Journal of Educational Psychology, 98*(3), 508-528.

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006b). Paraeducator-supplemented instruction in structural analysis with text reading practice for second and third graders at risk for reading problems. *Remedial and Special Education, 27*(6), 365-378. doi:10.1177/07419325060270060601

Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., . . . Francis, D. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *Elementary School Journal, 107*(2), 153-181. doi:10.1086/510653

Wang, C., & Algozzine, B. (2008). Effects of targeted intervention on early literacy skills of at-risk students. *Journal of Research in Childhood Education, 22*(4), 425. Retrieved from http://www.acei.org

Wanzek, J., & Vaughn, S. (2008). Response to varying amounts of time in reading intervention for students with low response to intervention. *Journal of Learning Disabilities, 41*(2), 126-142.

Wehby, J. H., Dodge, K. A., & Greenberg, M. (1993). *Classroom Atmosphere Rating Scale.* Unpublished technical manual, Nashville, TN: Vanderbilt University.

Wheeler, J. J., Baggett, B. A., Fox, J., & Blevins, L. (2006). Treatment integrity: A review of intervention studies conducted with children with autism. *Focus on Autism and Other Developmental Disabilities, 21*(1), 45.

Wheeler, J. J., Mayton, M. R., Carter, S. L., Chitiyo, M., Menendez, A. L., & Huang, A. (2009). An assessment of treatment integrity in behavioral intervention studies conducted with persons with mental retardation. *Education and Training in Developmental Disabilities, 44*(2), 187-195.

Zipoli, R. P., Jr., Coyne, M. D., & McCoach, D. B. (2011). Enhancing vocabulary intervention for kindergarten students: Strategic integration of semantically

related and embedded word review. *Remedial and Special Education, 32*(2), 131-143. doi:10.1177/0741932510361262

APPENDIX A

EXAMINATION RESULTS OF EARLY READING INTERVENTION STUDIES

Number of Dimensions	Number of Studies (%)	**Dimensions of Fidelity**					Study Reference Numbers (from Appendix B)
		Adherence	Dosage	Program Diff.	Quality	Student Part.	
1	27 (32.1)	■					1,2,7,8,9,10,12,13,20,21,27,29,31,40, 41 42, 47, 50, 54, 55, 57, 81, 82, 84
2	5 (6.0)	■	■				3,26,70,71,73
	1 (1.2)	■		■			5
	6 (7.1)	■			■		28, 36, 63, 66, 67, 80
	2 (2.4)		■		■		4, 16
3	3 (3.6)	■	■	■			35, 39, 61
	3 (3.6)	■	■		■		32, 68, 72
	1 (1.2)	■	■			■	33
	6 (7.1)	■			■	■	23, 46, 69, 76, 77, 79
4	1 (1.2)	■	■	■	■		51
	1 (1.2)	■	■	■		■	59
	3 (3.6)	■		■	■	■	45, 48, 49
5	1 (1.2)	■	■	■	■	■	22
None	24 (28.6)						
Totals:		58	17	10	22	12	

APPENDIX B

BIBLIOGRAPHY OF EARLY READING INTERVENTION STUDIES

Following is the bibliography of the 84 early reading intervention articles examined in this study and referenced in Appendix A.

1. Al Otaiba, S., Schatschneider, C., & Silverman, E. (2005). Tutor-assisted intensive learning strategies in kindergarten: How much is enough? *Exceptionality, 13*(4), 195-208.
2. Algozzine, B., Marr, M. B., Kavel, R. L., & Dugan, K. K. (2009). Using peer coaches to build oral reading fluency. *Journal of Education for Students Placed at Risk (JESPAR), 14*(3), 256-270.
3. Almaguer, I. (2005). Effects of dyad reading instruction on the reading achievement of Hispanic third-grade English language learners. *Bilingual Research Journal, 29*(3), 509-526.
4. Amendum, S. J., Vernon-Feagans, L., & Ginsberg, M. C. (2011). The effectiveness of a technologically facilitated classroom-based early reading intervention. *The Elementary School Journal, 112*(1), 107-131.
5. Apthorp, H. S. (2006). Effects of a supplemental vocabulary program in third-grade reading/language arts. *Journal of Educational Research, 100*(2), 67-79.
6. Bauserman, K. L., Cassady, J. C., & Smith, L. L. (2005). Kindergarten literacy achievement: The effects of the PLATO integrated learning system. *Reading Research and Instruction, 44*(4), 49-60.
7. Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *Elementary School Journal, 107*(3), 251-273.
8. Begeny, J. C. (2011). Effects of the Helping Early Literacy with Practice Strategies (HELPS) reading fluency program when implemented at different frequencies. *School Psychology Review, 40*(1), 149-157.
9. Begeny, J. C., Laugle, K. M., Krouse, H. E., Lynn, A. E., Tayrose, M. P., & Stage, S. A. (2010). A control-group comparison of two reading fluency programs: The Helping Early Literacy with Practice Strategies (HELPS) program and the Great Leaps K-2 reading program. *School Psychology Review, 39*(1), 137-155.
10. Borman, G. D., & Dowling, N. M. (2009). Student and teacher outcomes of the superkids quasi-experimental study. *Journal of Education for Students Placed at Risk (JESPAR), 14*(3), 207-225.

11. Borman, G. D., Slavin, R. E., Cheung, A. C. K., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of success for all. *American Educational Research Journal, 44*(3), 701-731.

12. Brabham, E. G., Murray, B. A., & Bowden, S. H. (2006). Reading alphabet books in kindergarten: Effects of instructional emphasis and media practice. *Journal of Research in Childhood Education, 20*(3), 219.

13. Calhoon, M. B., Al Otaiba, S., Greenberg, D., King, A., & Avalos, A. (2006). Improving reading skills in predominantly Hispanic Title 1 first-grade classrooms: The promise of peer-assisted learning strategies. *Learning Disabilities Research & Practice, 21*(4), 261-272. doi:10.1111/j.1540-5826.2006.00222.x

14. Case, L. P., Speece, D. L., Silverman, R., Ritchey, K. D., Schatschneider, C., Cooper, D. H., . . . Jacobs, D. (2010). Validation of a supplemental reading intervention for first-grade children. *Journal of Learning Disabilities, 43*(5), 402-417. doi.org/10.1177/0022219409355475

15. Cassady, J. C., & Smith, L. L. (2005). The impact of a structured integrated learning system on first-grade students' reading gains. *Reading & Writing Quarterly, 21*(4), 361-376.

16. Chambers, B., Cheung, A. C. K., Madden, N. A., Slavin, R. E., & Gifford, R. (2006). Achievement effects of embedded multimedia in a success for all reading program. *Journal of Educational Psychology, 98*(1), 232-237.

17. Chambers, B., Slavin, R. E., Madden, N. A., Abrami, P., Logan, M. K., & Gifford, R. (2011). Small-group, computer-assisted tutoring to improve reading outcomes for struggling first and second graders. *Elementary School Journal, 111*(4), 625-640.

18. Chambers, B., Slavin, R. E., Madden, N. A., Abrami, P. C., Tucker, B. J., Cheung, A., . . . Gifford, R. (2008). Technology infusion in success for all: Reading outcomes for first graders. *Elementary School Journal, 109*(1), 1-15.

19. Cole, J. M., & Hilliard, V. R. (2006). The effects of web-based reading curriculum on children's reading performance and motivation. *Journal of Educational Computing Research, 34*(4), 353-380.

20. Coyne, M. D., McCoach, D. B., & Kapp, S. (2007). Vocabulary intervention for kindergarten students: Comparing extended instruction to embedded instruction and incidental exposure. *Learning Disability Quarterly, 30*(2), 74-88.

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

22. Denton, C. A., Kethley, C., Nimon, K., Kurz, T. B., Mathes, P. G., Minyi, S., . . . Swanson, E. A. (2010). Effectiveness of a supplemental early reading intervention scaled up in multiple schools. *Exceptional Children, 76*(4), 394-416.

23. Denton, C. A., Solari, E. J., Ciancio, D. J., Hecht, S. A., & Swank, P. R. (2010). A pilot study of a kindergarten summer school reading program in high-poverty urban schools. *Elementary School Journal, 110*(4), 423-439.

24. DiLorenzo, K. E., Rody, C. A., Bucholz, J. L., & Brady, M. P. (2011). Teaching letter-sound connections with picture mnemonics: "Itchy's alphabet" and early decoding. *Preventing School Failure, 55*(1), 28-34.

25. Ehri, L. C., Dreyer, L. G., Flugman, B., & Gross, A. (2007). Reading rescue: An effective tutoring intervention model for language-minority students who are struggling readers in first grade. *American Educational Research Journal, 44*(2), 414-448.

26. Ehri, L. C., Satlow, E., & Gaskins, I. (2009). Grapho-phonemic enrichment strengthens keyword analogy instruction for struggling young readers. *Reading and Writing Quarterly, 25*(2-3), 162-191.

27. Fien, H., Santoro, L., Baker, S. K., Park, Y., Chard, D. J., Williams, S., . . . Haria, P. (2011). Enhancing teacher read alouds with small-group vocabulary instruction for students with low vocabulary in first-grade classrooms. *School Psychology Review, 40*(2), 307-318.

28. Frechtling, J. A., Zhang, X., & Silverstein, G. (2006). The voyager universal literacy system: Results from a study of kindergarten students in inner-city schools. *Journal of Education for Students Placed at Risk (JESPAR), 11*(1), 75-95.

29. Fuchs, D., Compton, D. L., Fuchs, L. S., Bryant, J., & Davis, G. N. (2008). Making "secondary intervention" work in a three-tier responsiveness-to-intervention model: Findings from the first-grade longitudinal reading study of the national research center on learning disabilities. *Reading and Writing: An Interdisciplinary Journal, 21*(4), 413-436.

30. Gest, S. D., & Gest, J. M. (2005). Reading tutoring for students at academic and behavioral risk: Effects on time-on-task in the classroom. *Education and Treatment of Children, 28*(1), 25-47.

31. Gunn, B., Smolkowski, K., Biglan, A., Black, C., & Blair, J. (2005). Fostering the development of reading skill through supplemental instruction: Results for Hispanic and non-Hispanic students. *Journal of Special Education, 39*(2), 66-85.

32. Gunn, B., Smolkowski, K., & Vadasy, P. (2011). Evaluating the effectiveness of "read well kindergarten." *Journal of Research on Educational Effectiveness, 4*(1), 53-86.

33. Hall, K. M., Sabey, B. L., & McClellan, M. (2005). Expository text comprehension: Helping primary-grade teachers use expository texts to full advantage. *Reading Psychology an International Quarterly, 26*(3), 211-234.

34. Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergartners. *Language, Speech, and Hearing Services in Schools, 36*(1), 17-32.

35. Kamps, D., Abbot, M., Greenwood, C., Arreaga-Mayer, C., Wills, H., Longstaff, J., . . . Walton, C. (2007). Use of evidence-based, small-group reading instruction for English language learners in elementary grades: Secondary-tier intervention. *Learning Disability Quarterly, 30*(3), 153-168.

36. Kamps, D. M., & Greenwood, C. R. (2005). Formulating secondary-level reading interventions. *Journal of Learning Disabilities, 38*(6), 500-509.

37. Kerins, M. R., Trotter, D., & Schoenbrodt, L. (2010). Effects of a Tier 2 intervention on literacy measures: Lessons learned. *Child Language Teaching and Therapy, 26*(3), 287-302.

38. Knezek, G., & Christensen, R. (2008). Effect of technology-based programs on first- and second-grade reading achievement. *Computers in the Schools, 24*(3-4), 23-41.

39. Kuhn, M. R., Schwanenflugel, P. J., Morris, R. D., Morrow, L. M., Woo, D. G., Meisinger, E. B., . . . Stahl, S. A. (2006). Teaching children to become fluent and automatic readers. *Journal of Literacy Research, 38*(4), 357-387.

40. Lane, H. B., Pullen, P. C., Hudson, R. F., & Konold, T. R. (2009). Identifying essential instructional components of literacy tutoring for struggling beginning readers. *Literacy Research and Instruction, 48*(4), 277-297.

41. Lewandowski, L., Begeny, J., & Rogers, C. (2006). Word-recognition training: Computer versus tutor. *Reading & Writing Quarterly, 22*(4), 395-410.

42. Lo, Y., Wang, C., & Haskell, S. (2009). Examining the impacts of early reading intervention on the growth rates in basic literacy skills of at-risk urban kindergarteners. *The Journal of Special Education, 43*(1), 12-28. doi:10.1177/0022466907313450

43. Macaruso, P., Hook, P. E., & McCabe, R. (2006). The efficacy of computer-based supplementary phonics programs for advancing reading skills in at-risk elementary students. *Journal of Research in Reading, 29*(2), 162-172.

44. Macaruso, P., & Walker, A. (2008). The efficacy of computer-assisted instruction for advancing literacy skills in kindergarten children. *Reading Psychology, 29*(3), 266-287.

45. Mathes, P. G., Denton, C. A., Fletcher, J. M., Anthony, J. L., Francis, D. J., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly, 40*(2), 148-182.

46. Maynard, K. L., Pullen, P. C., & Coyne, M. D. (2010). Teaching vocabulary to first-grade students through repeated shared storybook reading: A comparison of rich and

basic instruction to incidental exposure. *Literacy Research and Instruction, 49*(3), 209-242.

47. McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children, 71*(4), 445-463.
48. McMaster, K. L., Kung, S., Han, I., & Cao, M. (2008). Peer-assisted learning strategies: A "tier 1" approach to promoting English learners' response to intervention. *Exceptional Children, 74*(2), 194-214.
49. Nelson, J. R., Stage, S. A., Epstein, M. H., & Pierce, C. D. (2005). Effects of a prereading intervention on the literacy and social skills of children. *Exceptional Children, 72*(1), 29.
50. Nelson, J. R., Vadasy, P. F., & Sanders, E. A. (2011). Efficacy of a Tier 2 supplemental root word vocabulary and decoding intervention with kindergarten Spanish-speaking English learners. *Journal of Literacy Research, 43*(2), 184-211.
51. O'Connor, R. E., Bocian, K., Beebe-Frankenberger, M., & Linklater, D. L. (2010). Responsiveness of students with language difficulties to early intervention in reading. *Journal of Special Education, 43*(4), 220-235.
52. Osborn, J., Freeman, A., Burley, M., Wilson, R., Jones, E., & Rychener, S. (2007). Effect of tutoring on reading achievement for students with cognitive disabilities, specific learning disabilities, and students receiving title I services. *Education and Training in Developmental Disabilities, 42*(4), 467-474.
53. Paris, A. H., & Paris, S. G. (2007). Teaching narrative comprehension strategies to first graders. *Cognition and Instruction, 25*(1), 1-44.
54. Puhalla, E. M. (2010). Enhancing the vocabulary knowledge of first-grade children with supplemental booster instruction. *Remedial and Special Education*. Advance online publication. doi:10.1177/0741932510362495
55. Rafdal, B. H., McMaster, K. L., McConnell, S. R., Fuchs, D., & Fuchs, L. S. (2011). The effectiveness of kindergarten peer-assisted learning strategies for students with disabilities. *Exceptional Children, 77*(3), 299-316.
56. Reutzel, D. R., Fawson, P. C., & Smith, J. A. (2008). Reconsidering silent sustained reading: An exploratory study of scaffolded silent reading. *Journal of Educational Research, 102*(1), 37-50.
57. Reutzel, D. R., Smith, J. A., & Fawson, P. C. (2005). An evaluation of two approaches for teaching reading comprehension strategies in the primary years using science information texts. *Early Childhood Research Quarterly, 20*(3), 276-305.
58. Schacter, J., & Jo, B. (2005). Learning when school is not in session: A reading summer day-camp intervention to improve the achievement of exiting first-grade students who are economically disadvantaged. *Journal of Research in Reading, 28*(2), 158-169.

59. Schwanenflugel, P. J., Kuhn, M. R., Morris, R. D., Morrow, L. M., Meisinger, E. B., Woo, D. G., . . . Sevcik, R. (2009). Insights into fluency instruction: Short- and long-term effects of two reading programs. *Literacy Research and Instruction, 48*(4), 318-336.

60. Schwartz, R. M. (2005). Literacy learning of at-risk first-grade students in the reading recovery early intervention. *Journal of Educational Psychology, 97*(2), 257-267.

61. Silverman, R. (2007). A comparison of three methods of vocabulary instruction during read-alouds in kindergarten. *Elementary School Journal, 108*(2), 97-113.

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

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

64. Stevens, R. J., Van Meter, P. N., Garner, J., Warcholak, N., Bochna, C., & Hall, T. (2008). Reading and integrated literacy strategies (RAILS): An integrated approach to early reading. *Journal of Education for Students Placed at Risk (JESPAR), 13*(4), 357-380.

65. Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Herron, J., & Lindamood, P. (2010). Computer-assisted instruction to prevent early reading difficulties in students at risk for dyslexia: Outcomes from two instructional approaches. *Annals of Dyslexia, 60*(1), 40-56.

66. Vadasy, P. F., & Sanders, E. A. (2008). Code-oriented instruction for kindergarten students at risk for reading difficulties: A replication and comparison of instructional groupings. *Reading and Writing: An Interdisciplinary Journal, 21*(9), 929-963.

67. Vadasy, P. F., & Sanders, E. A. (2008). Repeated reading intervention: Outcomes and interactions with readers' skills and classroom instruction. *Journal of Educational Psychology, 100*(2), 272-290.

68. Vadasy, P. F., & Sanders, E. A. (2009). Supplemental fluency intervention and determinants of reading outcomes. *Scientific Studies of Reading, 13*(5), 383-425.

69. Vadasy, P. F., & Sanders, E. A. (2010). Efficacy of supplemental phonics-based instruction for low-skilled kindergarteners in the context of language minority status and classroom phonics instruction. *Journal of Educational Psychology, 102*(4), 786-803.

70. Vadasy, P. F., & Sanders, E. A. (2011). Efficacy of supplemental phonics-based instruction for low-skilled first graders: How language minority status and pretest

characteristics moderate treatment response. *Scientific Studies of Reading, 15*(6), 471-497. doi:10.1080/10888438.2010.501091

71. Vadasy, P. F., Sanders, E. A., & Abbott, R. D. (2008). Effects of supplemental early reading intervention at 2-year follow up: Reading skill growth patterns and predictors. *Scientific Studies of Reading, 12*(1), 51-89.

72. Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2005). Relative effectiveness of reading practice or word-level instruction in supplemental tutoring: How text matters. *Journal of Learning Disabilities, 38*(4), 364.

73. Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. *Journal of Educational Psychology, 98*(3), 508-528.

74. Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006). Paraeducator-supplemented instruction in structural analysis with text reading practice for second and third graders at risk for reading problems. *Remedial and Special Education, 27*(6), 365-378.

75. Vadasy, P. F., Sanders, E. A., & Tudor, S. (2007). Effectiveness of paraeducator-supplemented individual instruction: Beyond basic decoding skills. *Journal of Learning Disabilities, 40*(6), 508-525.

76. Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., . . . Francis, D. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *Elementary School Journal, 107*(2), 153-181.

77. Vaughn, S., Wanzek, J., Murray, C. S., Scammacca, N., Linan-Thompson, S., & Woodruff, A. L. (2009). Response to early reading intervention: Examining higher and lower responders. *Exceptional Children, 75*(2), 165-183.

78. Vellutino, F. R., Scanlon, D. M., Zhang, H., & Schatschneider, C. (2008). Using response to kindergarten and first grade intervention to identify children at-risk for long-term reading difficulties. *Reading and Writing: An Interdisciplinary Journal, 21*(4), 437-480.

79. Wang, C., & Algozzine, B. (2008). Effects of targeted intervention on early literacy skills of at-risk students. *Journal of Research in Childhood Education, 22*(4), 425.

80. Wanzek, J., & Vaughn, S. (2008). Response to varying amounts of time in reading intervention for students with low response to intervention. *Journal of Learning Disabilities, 41*(2), 126-142.

81. Williams, J. P., Nubla-Kung, A. M., Pollini, S., Stafford, K. B., Garcia, A., & Snyder, A. E. (2007). Teaching cause-effect text structure through social studies content to at-risk second graders. *Journal of Learning Disabilities, 40*(2), 111-120.

82. Williams, J. P., Stafford, K. B., Lauer, K. D., Hall, K. M., & Pollini, S. (2009). Embedding reading comprehension training in content-area instruction. *Journal of Educational Psychology, 101*(1), 1-20.

83. Zipke, M., Ehri, L. C., & Cairns, H. S. (2009). Using semantic ambiguity instruction to improve third graders' metalinguistic awareness and reading comprehension: An experimental study. *Reading Research Quarterly, 44*(3), 300-321.

84. Zipoli, R. P., Jr., Coyne, M. D., & McCoach, D. B. (2011). Enhancing vocabulary intervention for kindergarten students: Strategic integration of semantically related and embedded word review. *Remedial and Special Education, 32*(2), 131-143.

APPENDIX C

FIDELITY OBSERVATION GUIDE

Teacher Group ID _____ Lesson Date_____

Observation #_____ Reliability_____ Primary _____ Observer_____ Date Coded_____

Scale: 0 = Not observed (i.e., never)
1 = Observed with low quality low/inconsistent high
2 = Observed with adequate quality (75%)
3 = Observed consistently with exceptional quality (almost 100%; most or all)

Instructional Practices	0-3	
Models new skills and strategies during lesson		
Provides guided practice		
Uses manipulatives/hands-on opportunities during instruction		
Provides plentiful opportunities for students to respond		
Provides systematic instruction that follows an instructional sequence (activities build on each other)		
Teacher Responsiveness		
Uses appropriate pacing (fluent, "just right")		
Modifies and/or scaffolds instruction when necessary		
Provides immediate feedback to students		
Student Engagement		
Students are engaged		
Students refrain from problem behavior		
	/33	%

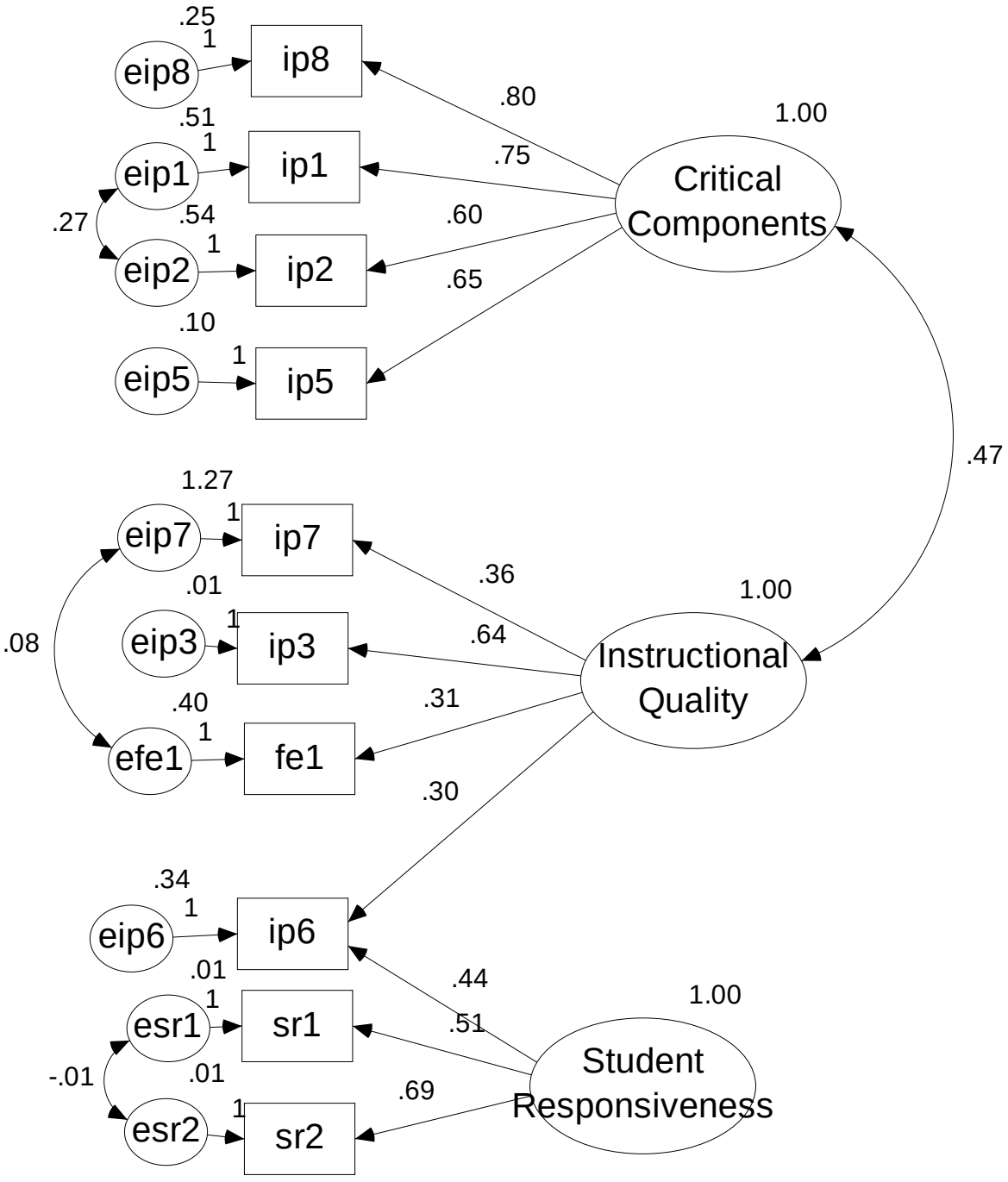
APPENDIX D

FIDELITY OBSERVATION GUIDE CODEBOOK

Fidelity Observation Guide Codebook
Instructional Practices
Models new skills and strategies during lesson • **Examples:** Teacher describes and demonstrates task before students begin activity. • **Non-examples**: Teacher does not describe and demonstrate task before students begin activity; may look like an assessment or review.
Provides guided practice (lead) • **Examples**: Teacher and students perform task together. Teacher guides the students through the task in order to achieve success. Group and/or individually. • **Non-examples:** Teacher and students do not perform task together. Teacher does not provide guidance as students work through the task.
Provides systematic instruction that follows an instructional sequence (activities build on each other) • **Examples:** Instruction follows a sequence and activities build on each other. The level of difficulty follows a continuum, for example, segmenting orally, segment, word building, reading connected text. • **Non-examples:** Instruction does not follow a sequence. The activities seem random and do not connect to each other. A continuum of difficulty is not evident.
Uses manipulatives/hands-on opportunities during instruction • **Examples**: Teacher uses hands-on activities such as picture cards, letter cards, and say-it, move-it boards. • **Non-examples:** Teacher does not use hands-on activities but rather talks about concepts and skills without manipulatives.
Teacher Responsiveness
Uses appropriate pacing (fluent, "just right") • **Examples:** Teacher's pacing is "just right." Teacher keeps a quick pace when students are responding correctly and slows down when needed to provide extra support or review. • **Non-examples:** Teacher's pacing is too fast when students need extra review or support. Students may seem lost or frustrated due to the fast pacing. Alternatively, teacher's pacing is too slow when students are able to move on. Students may seem distracted or bored due to the slow pacing.
Modifies and/or scaffolds instruction when necessary • **Examples:** Teacher may modify the content or level of difficulty of task, and/or teacher may scaffold the current content through physical or verbal prompting/support. • **Non-examples:** Teacher provides all students with the same task without support regardless of student need. All children receive the same task, etc.
Provides immediate feedback to students • **Examples:** Teacher provides feedback to student immediately after student response, for example, "Yes, good job." **Non-examples**: Teacher does not provide immediate feedback to student after a response. For example, the student answers and the teacher moves on.
Student Engagement
Students are engaged • **Examples:** Students are actively participating in activities. Attention and eyes are focused on the teacher and/or materials and the student is actively reading and writing. • **Non-examples**: Students are not participating in activities. Students may be playing with manipulatives rather than using them for instructional purposes. Students may also appear to be "zoning out" or talking off topic with peers, etc.
Students refrain from problem behavior • **Examples:** Students are not demonstrating behavior that interferes with learning. Students do not require redirection. • **Non-examples**: Students need constant redirection. Student behavior interferes with others learning.

APPENDIX E

CONFIRMATORY FACTOR ANALYSIS FOR OBSERVATION 2 AND 3



Variables: ip8—uses manipulatives; ip1—models new skills; ip2—guided practice; ip5—systematic instruction; ip7—modifies instruction; ip3—opportunities to respond; ip6—appropriate pacing; fe1—immediate feedback; sr1—student engagement; sr2—students refrain from problems.

Using the results from the exploratory factor analysis from Observation 1, the same three-factor model was tested using confirmatory factor for both Observation 2 and 3. For Observation 2, the pattern coefficients (factor loadings) were allowed to be estimated. However, for Observation 3, the pattern coefficients were constrained to match the pattern coefficients of Observation 2. Therefore, there is only one model represented in Appendix E. The difference will be in the model fit test statistics.

The critical component factor had the four observed variables from the EFA loaded on it. The instructional quality and student responsive factors were each indicated by three observed variables. All three of the factor parameters were constrained to equal 1. Initially, all factors were initially correlated and there were no residual variance correlated. This model did not fit well statistically.

Modification indices were consulted. The residual correlations between certain variables—(a) models new skills and provides guided practice, (b) modifies instruction when necessary and provides immediate feedback, and (c) students are engaged and students refrain from problem behavior—were added based on the modification indices suggestions. Also, the correlation between the student responsiveness factor and the other factors was removed.

All standardized pattern coefficients (regression weights) were statistically significant for all factors with the exception of the variable modifies instruction. The values of the coefficients ranged from .301 to .799 for observation.

The following fit indices were employed: (a) the Comparative Fit Index (CFI; Bentler, 1990), with values greater than .95 indicating reasonable model fit; and (b) the Root Mean Square Error of Approximation (RMSEA; Steiger, 1990), with values less than .08 indicating reasonable model fit. A model was determined to fit well if both criteria were met. The Observation 2 model fit was (χ^2 [33, N = 23] = 41.97, p = .136. The CFI = .921 and the RMSEA = .114 indicated a poor fit. The Observation 3 model fit was (χ^2 [45, N = 23] = 61.74, p = .049. The CFI = .839 and the RMSEA = .130 indicated a poor fit.

VITA

Melissa Shea Fogarty received her Bachelor of Science degree in applied learning and development from The University of Texas at Austin in 2002. She received her Master of Education degree in special education in 2006 from Texas State University. Her research interests include special education, struggling readers, fidelity of implementation, and teacher quality.

Mrs. Fogarty may be reached at 4203 Cedar Creek Court, College Station, Texas, 77845. Her email is Melissafogarty@gmail.com