August 2015

Examining Whether School Finances and Academic Achievement Predict the Quality of Reading Intervention Implementation in Response-to-Intervention

Joshua Anderw Looser
University of Wisconsin-Milwaukee

EXAMINING WHETHER SCHOOL FINANCES AND ACADEMIC ACHIEVEMENT PREDICT THE QUALITY OF READING INTERVENTION IMPLEMENTATION IN RESPONSE-TO-INTERVENTION

by

Joshua A. Looser

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Educational Psychology

at

The University of Wisconsin-Milwaukee

August 2015

ABSTRACT

EXAMINING WHETHER SCHOOL FINANCES AND ACADEMIC ACHIEVEMENT PREDICT THE QUALITY OF READING INTERVENTION IMPLEMENTATION IN RESPONSE-TO-INTERVENTION

by

Joshua A. Looser

The University of Wisconsin-Milwaukee, 2015
Under the Supervision of Professor Dr. Markeda Newell

The purpose of this study was to examine how financial capacity and reading achievement contribute to the implementation of high-quality reading interventions in the context of RtI. As a secondary research interest, the relationship between reading achievement and intervention intensity was examined. Financial capacity was operationalized in terms of per-pupil expenditure, while achievement was examined based on performance on the state standardized test. The quality of reading interventions was defined by four indicators: 1) evidence-based reading interventions, 2) psychometrically sound progress monitoring, 3) treatment integrity measures, and 4) interventionist training. Data regarding these four indicators and intervention intensity were obtained via a state administered survey (92 respondents with a response rate of 59.0%). Using a regression-based approach (i.e., linear and ordinal regression), the present findings indicated that reading proficiency and per-pupil-expenditure were not significantly predictive of the use of high-quality reading interventions at Tier 2 or Tier 3, across any of the four quality indicators: evidence-based interventions, progress

monitoring tools, treatment integrity measures, and interventionist training. The lack of significant relationships between per-pupil expenditure, achievement levels, and the quality of evidence-based reading interventions is hypothesized to have occurred for two inter-related reasons: 1) schools appear to be at varying stages of RtI implementation, and 2) most schools appear to be at an early stage of implementation. Consistent with expectations, and the theoretical RtI model, intervention provision at Tier 3 was significantly more intense than at Tier 2. This finding is encouraging in terms of schools personnel's' capacity to provide interventions at varying intensities based on student need. Despite these expected findings, a significant relationship was not observed between achievement levels and intervention intensity at Tier 2 or Tier 3. Collectively, results indicated that the impact of school-level funding and school-level achievement may be too distally related to the provision of interventions at an early point in implementation. These results highlight the necessity for conceptualizing school-based program implementation from a theoretical perspective, which will enable an understanding of how systems-level variables differentially impact implementation across stages and ensure that schools can be appropriately supported in their implementation.

ACKNOWLEDGEMENTS

I would first like to thank my dissertation committee, Drs. Markeda Newell, Karen Stoiber, David Klingbeil, Nancy Rice, and Cindy Walker, for their time and support throughout this process. I would like to give a special thank you to Dr. Markeda Newell, for both her support with this project and also throughout my graduate schooling. I still remember the day when I came to UW-Milwaukee as a prospective graduate student to learn more about the program (with my dad tagging along). In the approximately 5 years since then, I now attribute my passion, ideological development, and professional growth to have been a direct result of working with you and under your tutelage. I cannot thank you enough and would not have reached this milestone without your continued support.

Lastly, I would like to thank my family (dad, mom, and sister) for making me the person I am today. You have personified the meaning of father, mother, and sibling; you have continually served to motivate me and challenge me to become the best version of myself that I can be, both professionally and personally. Of course, the non-interest loans have also been critical! Thank you for everything.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER ONE

Introduction

The evidence-based practice movement is based on the identification, dissemination and adoption of practices that have underlying empirical support (Kratochwill, 2007). This movement is especially important in reading considering the large percentage of children at risk for reading failure in the United States (32% are below basic in fourth grade; National Center for Research Statistics, 2013). The poor reading skill of these children is, in part, indicative of an insufficient early instructional environment (Foorman & Torgesen, 2001). To address the challenges of students at risk for reading failure, systems of preventative, evidence-based interventions have been recommended (American Psychological Association, 2005; National Association of School Psychologists, 2007; National Joint Committee on Learning Disabilities, 2005). The use of evidence-based interventions is recommended because they have been empirically shown to address the needs of many students at risk for reading difficulties (Chard, Stoolmiller, Harn, Wanzek, Vaughn, Linan-Thompson, & Kame'enui, 2008; Goss & Brown-Chidsey, 2012; VanDerHeyden, Witt, Gilbertson, 2007; Snyder, Broussard, & Ramsdell, 2007; Wanzek & Vaughn, 2008). Specifically, a model known as response-to-intervention (RtI) has risen in prominence and embodies the preventative and evidence-based intent of the movement.

RtI stems from the work of Deno's (1985) data-based program modification model and Bergan's behavioral consultation models (representing the foundations of the problem-solving framework) (Bergan, 1977; Bergan & Kratochwill, 1990), and is also associated with recent legislation, including the Individuals with Disabilities Education

Improvement Act (2004) and No Child Left Behind (2001). The RtI model is commonly conceptualized as a 3-tiered model of service delivery (Reschly, 2008). Tier 1 involves services provided to all students and is intended to be effective for at least 80% of students. Universal screening is also part of Tier 1 and serves to ensure that the general education curriculum is effective for students (at least 80%), while also allowing the identification of students who may be at risk for reading failure (Batsche, Elliot, Graden, Grimes, Kovaleski, Prasse, et al., 2005; Wanzek & Vaughn, 2007). Tier 2 is reserved for students who remain below screening benchmarks and require additional support in the form of interventions. Typically, the at-risk students are provided strategic, small-group interventions that have underlying empirical evidence to support their use (Brown-Chidsey & Steege, 2010). Tier 2 interventions (plus Tier 1) are conceptualized to be effective for approximately 15% of students. Intervention effectiveness (i.e., student responsiveness) is determined based on the results of ongoing progress monitoring; conversely, progress monitoring data is used to determine if a student is non-respondent and requires services beyond Tier 2 (Fuchs & Fuchs, 2002). Additional services provided to students who are non-respondent to Tier 2 are known as Tier 3 services, which are distinguished by an increase in intensity and greater individualization of intervention. Therefore, Tier 3 services involve more intensive, long-term interventions, with smaller groups of students, and more frequent progress monitoring (Institute for Education Sciences, 2009; Wanzek & Vaughn, 2007).

Given the emphasis on preventative evidence-based interventions within an RtI model, it aligns well with the evidence-based practice movement by providing an avenue for the delivery of evidence-based interventions in school-based contexts (Denton, Tolar,

Fletcher, Barth, Vaughn, & Francis, 2013). The implementation of interventions is a substantial requirement for schools because they are essential to providing students with effective services (consistent with the aims of the evidence-based practice movement) (Gansle & Noell, 2007). However, surprisingly, the implementation of interventions has commonly been assumed rather than considered a major issue to be resolved (Gansle & Noell, 2007).

The lack of study on intervention implementation in the context of RtI relates to a more general phenomenon known as the 'research-to-practice' gap. The gap is characterized by a rift between the research and practice worlds. In the realm of research, many programs have been developed and shown to be effective, and yet, they fail to cross-over into the realm of practice, despite considerable diffusion and dissemination strategies (Ogden & Fixsen, 2014). The growing awareness of this gap has led to research efforts focusing on implementation facilitators and obstacles, and led to the emergence of the field known as implementation science (Institute of Medicine, 2001). Implementation science takes an empirical approach to the systematic uptake of research findings into routine practice. In many ways it addresses an irony within the evidence-based practice movement, where the implementation of evidence-based interventions is promoted without an evidence-base to guide the implementation process (Ogden & Fixsen, 2014).

Understanding the barriers and facilitators to implementation are critically important in the context of RtI because the model relies on the uptake of the empirically-based interventions in the context of routine practice. Therefore, the implementation of evidence-based reading interventions in the context of RtI represents an important

outcome of empirical interest. Specifically, the following four components are indicators of quality reading intervention implementation in the context of RtI: 1) the use of an evidence-based reading intervention, 2) the use of psychometrically sound progress monitoring, 3) the use of treatment integrity measures, and 4) the use of a trained interventionist.

Components of Evidence-Based Reading Intervention Implementation

Evidence-based intervention. A core element of evidence-based practice is the use of evidence-based interventions, which are interventions based on the best available research that demonstrates their efficacy and/or effectiveness (APA Presidential Task Force on Evidence-Based Practice, 2006). Evidence-based interventions are based on research findings that demonstrate they are likely to produce predictable, beneficial, and effective results (Forman & Burke, 2008). There is considerable support for the use of Tier 2 reading interventions (Chard et al., 2008; Goss & Brown-Chidsey, 2012; VanDerHeyden et al., 2007) and emerging support for the use of Tier 3 reading interventions (Al Otaiba, Connor, Folsom, Greulich, Meadows, & Li, 2011; Conner, Morrison, Fishman, Schatschneider, & Underwood, 2007). Despite the empirical foundation of reading interventions, research can include many designs and there are varying standards for intervention quality (Stoiber & Desmet, 2010). Considering the complexities of determining what is 'evidence-based', multiple groups have begun reviewing academic and behavioral programs to support implementation in the schools by providing evidence ratings on the level of research underlying a variety of interventions (Slavin, 2008). While many groups focus on mental health, the following groups review reading interventions in an ongoing fashion: What Works Clearinghouse,

Promising Practices Network, and Johns Hopkins Best Evidence Encyclopedia. Since these groups utilize different methodologies for review and standards of quality, using a combination of these databases represents a robust method for determining interventions that are 'evidence-based'. While the use of an evidence-based intervention is a critical foundation of the implementation of high-quality reading interventions, progress monitoring is necessary for evaluating and improving program effects (Deno, 2003; Deno, 2005).

Progress monitoring. Progress monitoring is when teachers (or other personnel) assess students' academic performance on a regular basis (Fuchs & Fuchs, 2002). It allows the documentation of student growth as well as a determination of whether children are appropriately benefiting from the given instructional program (i.e., evidence-based intervention). Curriculum-based measurement (CBM) is the preferred tool for this function (Fuchs & Fuchs, 2002). CBM is a set of standardized and short duration tests that can be used by educators to evaluate the effects of instructional interventions (Shinn, 2008). Curriculum-based measures are preferred relative to other methods because they have: 1) documented reliability and validity for multiple forms (e.g. oral reading fluency, maze, and word identification) (Marston, 1989; Wayman, Wallace, Wiley, Ticha, & Espin, 2007), 2) been shown to predict important outcomes (e.g. state standardized test results) (Good, Simmons, & Kame'enui, 2001; Wayman et al., 2007), and 3) emerging support for their use in the context of instructional decision-making (Christ, Zopluoglu, Monaghen, & Van Norman, 2013; Institute for Education Sciences, 2009).

In addition to these three characteristics, it is critical that progress monitoring occur sufficiently frequent. Recent evidence suggests that there are frequency thresholds

that promote reliable and valid slopes of progress for decision-making purposes (Christ et al., 2013). Specifically, it is recommended that progress monitoring occur 2 times per week at Tier 2 and 5 times per week at Tier 3 (Christ et al., 2013). Progress monitoring is a data collection tool that allows a determination of the effectiveness of the intervention, but it is also necessary to ensure treatment integrity in order to attribute effectiveness to the intervention itself.

Treatment integrity. Treatment integrity is defined as the degree to which an intervention is implemented as planned (Hagermoser Sanetti & Kratochwill, 2005). The purpose of assessing treatment integrity is to determine whether the intervention is being implemented appropriately, which ensures that accurate conclusions can be made regarding its effectiveness. Not only is treatment integrity significant for this interpretive purpose, it is also critical because of research has demonstrated that higher levels of treatment integrity results in better treatment outcomes (Durlak & DuPre, 2008; O'Donnel, 2008). Because of this relationship, treatment integrity is a necessity in the context of intervention implementation; however, training interventionists represents an additional component known to contribute to positive treatment outcomes.

Trained interventionist. Studies of preventive interventions comparing teachers who received in-service training to those without demonstrate that training, in general, is an important element for effective implementation (Perry, Murray, & Griffin, 1990; Ross, Leupker, Nelson, Saavedra, & Hubbard, 1991). Training has additional benefits beyond this general relationship. It has been shown that training contributes to the successful transportability and sustainability, such that the intervention is successfully 'transported' from the research realm, and sustained beyond original implementation (Forman, Serene

Olin, Eaton Hoagwood, Crowe, & Saka, 2009; Shernoff & Kratochwill, 2007). Additionally, training with performance feedback has been associated with greater integrity, which as described above, contributes to stronger outcomes and allows appropriate conclusions to be made (Noell, Witt, Gilbertson, Ranier, & Freeland, 1997; Noell, Witt, Slider, Connell, Gatti, et al., 2005; Witt, Noell, LaFleur, & Mortenson, 1997).

Together, these four components represent the basis of high-quality reading interventions in the context of RtI. They comprise a variable of interest, the 'what,' of high-quality reading intervention implementation in the context of RtI. However, in light of the significant research-to-practice gap, it is important to examine factors that contribute to successful implementation and better understand the 'how,' of high-quality reading interventions.

School Finances

High-quality evidence-based reading interventions can be costly. School personnel must select, likely purchase, and implement the interventions, develop progress monitoring procedures, develop integrity measurement procedures, and provide considerable training to school personnel to serve these functions. Simply put, the implementation of high-quality evidence-based interventions may cost considerable resources in terms of money and staff. Given the interest in examining barriers/facilitators to implementation, it is necessary to examine how schools are funded to determine if funding may play a potential role in enabling or hindering the implementation of high-quality reading interventions in the context of RtI.

Since education is primarily a state responsibility, more than 90% of school funding comes from state and local sources (Epstein, 2011). The funding system relies, largely, on local property taxes, which enables districts in high-wealth parts of a state to be funded more generously than districts in low-wealth areas; people living in property-rich districts can fund their public schools more liberally even at lower tax rates relative to residents in lower-income areas (Epstein, 2011). While there has been some reform to address this inequity, many states remain over-reliant on local property taxes and continue to fund their schools in ways that have no demonstrable link to the cost of delivering rigorous academic standards, particularly with children at risk for difficulties (Baker, Sciarra, and Farrie, 2014).

When examining school funding in a given state, initial considerations include the amount of funding that is received, and the sources of the funding. However, there are two additional measures of funding that contribute to the understanding of whether funding promotes equitable educational opportunities. First, there is funding distribution, which is based on the predicted per-pupil funding across levels of poverty (Baker et al., 2014). Funding distributions can be progressive, flat, or regressive. A progressive system is when states provide additional funds to students of higher poverty levels to address their additional needs, while a regressive system does the opposite (i.e., provides less funding for students of high poverty). A flat system provides equal funding independent of poverty level (Baker et al., 2014). Levels of poverty are significant because in order to promote equitable outcomes, funding levels must account for additional needs generated by student poverty (Baker et al., 2014). The second measure of funding is coverage, which is derived from the number of students within the public school system and the

degree of disparity between those within and outside the public school systems (Baker et al., 2014). Coverage is significant for two primary reasons. First, a higher number of students in the public system leads to a greater effort on the part of the state to fund those schools. Second, when students outside of the public sphere are from families with greater wealth, a significant source of funding is not made available to public schools (which is compounded if a large number of students are outside of the public school system) (Baker et al., 2014).

Within the state of study, the funding distribution and coverage indicate that the state is in the 'middle-of-the-pack' in terms of how the funding system promotes equity (Baker et al., 2014). Specifically, they utilize a flat funding distribution and have a relatively significant amount of students outside the public school system who, on average, are from more affluent families than those within the public system. Essentially, the coverage indicator suggests the pool of students attending public schools is at a slight disadvantage prior to consideration of the funding distribution; this is problematic when considering the state of interest provides the same amount of money for students across poverty levels, because the remainder of per-pupil expenditure dollars will come predominantly from local sources, serving to maintain the inequitable funding pattern that has been historically problematic (over-reliance on local funding) (Baker et al., 2014).

The faulty system of school funding highlights a potentially glaring challenge for the provision of high-quality reading interventions in schools. Reading interventions are part of the larger legislative aims of No Child Left Behind and the Individuals with Disabilities Education Improvement Act (2004), which target students at risk for academic difficulties, and yet, schools in lower-income areas receive less funding than

those in wealthier areas due to the dependence on local funding sources. This discrepancy engenders a contradiction where schools with the greatest need for the provision of high-quality evidence-based reading interventions may not be provided with the financial support to do so, leaving the intervention movement doomed to replicate the larger schooling inequities that have faced our nation for decades.

Achievement Levels of a Student Population

In addition to financial resources, the achievement levels of a student population can impact school personnel's capacity to implement high-quality reading interventions. The impact may occur indirectly and directly. The indirect impact of achievement levels can take place by creating a greater financial strain on schools. For example, the RtI model can be conceptualized as a resource-allocation model such that 15% of students are at risk for reading failure and receive Tier 2 supports, while 5% demonstrate significant difficulties in reading and need more intensive Tier 3 supports. However, achievement levels do not necessarily fit this template and low achieving schools will require greater amounts of each intervention implementation component; for example, they will need more personnel to implement the interventions, collect progress monitoring data, and collect treatment integrity data, they will need more intervention materials and progress monitoring probes, and they will need more personnel to be trained. In this manner, achievement levels can create additional financial strain, which as described above, is hypothesized to contribute to the ability to implement high-quality reading interventions.

Achievement levels may also directly impact the ability to implement interventions in adherence to the recommended practices of intervention implementation.

Specifically, low achieving schools will have more students in need of tiered supports (i.e. intervention and progress monitoring). Having large proportions of students in need of additional supports may create logistical challenges in providing interventions at the recommended dosage and group size (4 days per week, approximately 30 minutes per session, with groups of 4-6 at Tier 2; 5 days per week, with greater than 30 minutes per session and groups of 1-3 at Tier 3; Brown-Chidsey & Steege, 2010). In short, low achieving schools (i.e., schools with larger numbers of students needing support) may have difficulty in providing interventions of sufficient intensity.

Summary and Significance of Study

Response-to-intervention represents a potential mechanism to promote the use of evidence-based reading interventions in school-based contexts. Specifically, the use of evidence-based interventions, progress monitoring, treatment integrity and a trained interventionist represent the basis for implementation. The promotion of preventative reading interventions is particularly significant given the large numbers of students at risk for reading difficulties (National Center for Research Statistics, 2013). Its significance is compounded by the existing empirical knowledge of how to intervene with these students and the lack of empirical knowledge about how to translate the principles from research to practice settings. Accordingly, it is necessary to examine how to promote the implementation of high-quality reading interventions in practice settings so that students are provided with equitable educational opportunities.

When examining the process of school funding, it is possible that the current system does not enable the provision of high-quality reading interventions, particularly for students most in need. Limited coverage and a flat funding distribution lead to a

schooling context where the students who have the greatest need receive comparatively lesser funding, leaving the schools with limited capacity to preventatively remediate the difficulties through high-quality reading interventions. The problem is potentially compounded among low achieving schools in two ways. First, low achieving schools will have higher fiscal need in the form of greater intervention resources (e.g. amount of resources and amount personnel requiring training). Second, low achievement may create logistical challenges in adherence to recommended practices in the context of RtI (e.g. recommended group sizes, intervention duration, and frequency of progress monitoring). In short, we are faced with a harsh paradox in our schools that leaves legislation (NCLB and IDEA) and prominent educational movements (RtI and EBP) falling short of their intended goal: equitable and adequate educational opportunities for all students. Accordingly, this topic of study has implications for some of the most longstanding and persistent challenges in education.

Purpose and Research Questions

Based on the above rationale, the purpose of this study is to examine how financial capacity and achievement levels contribute to the implementation of high-quality reading interventions in the context of RtI. Specifically:

1. Does per-pupil expenditure and student's reading proficiency predict the quality of evidence-based interventions, progress monitoring tools, treatment integrity measures, and trained interventionists at Tier 2?

    1a. Does per-pupil expenditure and student's reading proficiency predict the use of quality of evidence-based interventions at Tier 2?

1b. Does per-pupil expenditure and student's reading proficiency predict the use of quality progress monitoring tools at Tier 2?

1c. Does per-pupil expenditure and student's reading proficiency predict the use of quality treatment integrity measures at Tier 2?

1d. Does per-pupil expenditure and student's reading proficiency predict the use of quality interventionist training at Tier 2?

2. Does per-pupil expenditure and student's reading proficiency predict the quality of evidence-based interventions, progress monitoring tools, treatment integrity measures, and interventionist training at Tier 3?

    2a. Does per-pupil expenditure and student's reading proficiency predict the use of quality of evidence-based interventions at Tier 3?

    2b. Does per-pupil expenditure and student's reading proficiency predict the use of quality progress monitoring tools at Tier 3?

    2c. Does per-pupil expenditure and student's reading proficiency predict the use of quality treatment integrity measures at Tier 3?

    2d. Does per-pupil expenditure and student's reading proficiency predict the use of quality interventionist training at Tier 3?

3. Does student's reading proficiency predict the intensity (based on dosage and group size) of intervention implementation at Tier 2?
4. Does student's reading proficiency predict the intensity (based on dosage and group size) of intervention implementation at Tier 3?

CHAPTER TWO

Literature Review

The American Psychological Association (APA) defined evidence-based practice in psychology as, "the integration of best available research with clinical expertise in the context of patient characteristics, culture, and preferences" (APA Presidential Task Force on Evidence-Based Practice, 2006, p. 273). While the use of science to inform practice is not new, there has been a renewed focus on the utilization of evidence-based interventions (EBI), which has become known as the evidence-based practice movement (APA Presidential Task Force on Evidence-Based Practice, 2006). The use of evidence-based interventions in schools is critically important in reading, especially when considering the large percentage of children at risk for reading failure in the United States (National Center for Research Statistics, 2013).

The fourth grade National Assessment of Educational Progress (NAEP) reading assessment shows that 32% of children are below a basic level in fourth grade (National Center for Research Statistics, 2013). Below basic is an achievement level that signifies the student is significantly behind grade-level benchmarks and has not achieved prerequisite knowledge and skills necessary for proficiency (National Center for Research Statistics, 2013). The poor reading skill of these children does not signify that 32% of children in the US have a reading disability, but that the instructional environment in early elementary school is insufficient (Foorman & Torgesen, 2001). Some students enter school with significant skill deficits and a lack of preparation, such that they require instructional interventions beyond the capacity of the typical classroom context (Foorman & Torgesen, 2001). However, other students may enter with adequate

skills, but have difficulty making adequate progress due to poor teaching (Fuchs & Fuchs, 2006). These factors indicate that risk for reading failure involves an interaction between child characteristics and the instructional context.

In response to the challenges of students at risk for reading failure, approaches that promote preventive, empirically-based services have been recommended (American Psychological Association, 2005; National Association of School Psychologists, 2007; National Joint Committee on Learning Disabilities, 2005). The use of empirically-based interventions beyond core instruction can address the needs of many students at risk for reading difficulties independent of the cause of difficulty (i.e., inadequate instruction or specific skill deficits / lack of preparation). Specifically, evidence-based interventions have been shown to immediately benefit students with risk for reading difficulties, and also prevent more significant reading problems (Chard et al., 2008; Goss & Brown-Chidsey, 2012; VanDerHeyden et al., 2007; Wanzek & Vaughn, 2008). An increasingly prominent approach to promote the use of preventative evidence-based reading practices in schools is through response-to-intervention (RtI), a movement in both general and special education of considerable interest to the profession of psychology (Kratochwill, Clements, Kalymon, 2007).

Evidence-Based Practice in Response-to-Intervention

The RtI model has risen in stature due to its connection with recent legislative mandates (Individuals with Disabilities Education Improvement Act, 2004; No Child Left Behind, 2001). Specifically, both IDEA and NCLB support the preventative intent of RtI and advocate for early identification of struggling students, while the use of response to scientific research-based interventions is explicitly cited in IDEA (in the evaluation of

specific learning disabilities, which represents the largest special education category (Simmons, Kame'enui, Coyne, & Chard, 2002)). RtI is not directly specified in the legislation, but it represents a prominent model that aligns with the use of scientifically-based interventions as stated in IDEA and NCLB. Further, based on a review of state laws, Zirkel (2012) identified that RtI has been specifically required by 12 states as of March, 2012, and there are many features of RtI that are required or recommended in other states (e.g. 10 states allow districts to develop RtI implementation plans and approximately half of the states require student intervention plans). With its legislative connection, schools have a legal impetus to implement the RtI model, and accordingly, it provides a potential mechanism to enable the use of evidence-based reading interventions in typical practice.

Response-to-intervention is the practice of using evidence-based instruction/intervention to address student needs while monitoring student progress over time in learning and/or behavioral domains (Brown-Chidsey & Steege, 2010). The foundations of the model stem from the work of Deno's (1985) data-based program modification model and Bergan's behavioral consultation models (i.e., representing the foundations of the problem-solving framework) (Bergan, 1977; Bergan & Kratochwill, 1990). RtI is most commonly described as a 3-tiered model of service delivery. Tier 1 refers to services provided to all students, which should be effective for approximately 80% of students (Batsche et al, 2005; Brown-Chidsey & Steege, 2010). Universal screening for all children ensures that the general education curriculum is functioning appropriately, while also serving the preventative function of identifying those who may be at risk. Essentially, Tier 1 is the systematic integration of data-based assessment

methods and high-quality teaching for all students (Brown-Chidsey & Steege, 2010; Wanzek & Vaughn, 2007).

Tier 2 is for those students who, despite sound Tier 1 curriculum/instruction, remain below benchmarks and thus require additional interventions. This tier is defined by the administration of strategic, small-group evidence-based interventions with integrity, and coupled with progress monitoring to examine their effectiveness. As conceptualized in most models, core (Tier 1) plus Tier 2 interventions are effective for 15% of students (Goss & Brown-Chidsey, 2012). Students deemed non-respondent based on progress monitoring data are provided Tier 3 services. Tier 3 is similar to Tier 2 in that it involves the administration of evidence-based interventions with integrity and progress monitoring. However, the distinction between Tier 2 and Tier 3 is that Tier 3 interventions are predicated on an increase in intensity and individualization of interventions, conceptualized to be necessary for approximately 5% of students. Tier 3 interventions involve more frequent progress monitoring and more intensive, long-term interventions, with smaller groups of students (Institute for Education Sciences, 2009; Wanzek & Vaughn, 2007).

Tiers 2 and 3 represent an attempt to match student need to services through the provision of evidence-based interventions (Brown-Chidsey & Steege, 2010). Accordingly, the significance of RtI in the context of the evidence-based practice movement in reading is that Tier 2 and Tier 3 interventions provide an avenue for the provision of evidence-based interventions in school-based contexts (Denton et al., 2013). Despite being a foundational element for successful implementation of RtI, the implementation of interventions has commonly been assumed rather than considered a

major issue to be resolved (Gansle & Noell, 2007). In fact, the implementation of interventions is a substantial requirement for schools, particularly when considering their provision is necessary to ensure that students are provided effective services (consistent with the aims of the evidence-based practice movement) (Gansle & Noell, 2007).

Implementation Science as a Framework for Evidence-Based Interventions in RtI

Even with the strong support of the evidence-based practice movement, the question of how to translate it effectively into practice remains substantial (Ogden & Fixsen, 2014). It was previously thought that if a practice or program was efficacious and if information were made available, it would be implemented. However, this is clearly not the case, diffusion and dissemination strategies have resulted in a 14% use of evidence-based programs after 17 years (Balas & Boren, 2000; Brownson, Colditz, & Proctor, 2012; Green, 2008). This phenomenon has become known as the 'research-to-practice' or 'science-to- service' gap because despite the availability of research-based practices, they are not typically implemented in practice contexts.

The low levels of implementation highlight a critical need for research. Diffusion and dissemination are not sufficient to reliably produce and sustain positive benefits to consumers (Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005). It is necessary to examine mechanisms that can support the adoption of evidence-based practices. The underlying science of this process is known as implementation science, which is defined as, "the scientific study of methods to promote the systematic uptake of clinical research findings and other evidence-based practices into routine practice..." (ICE-BeRG, 2006). In other words, it is the study of the processes or change mechanisms that bring about successful implementation of evidence-based interventions. Research has begun to

examine the process of practice change, which, in general, is associated with the characteristics of the innovation itself, the provider, the practitioner adopting the practice, the client or consumer, and the inner and outer context of the service delivery organization (Durlak & DuPre, 2008; Greenhalgh, Robert, MacFarlane, Bate, & Kyriakidou, 2004). Given the challenges of implementation and the strong impetus underlying RtI, it is critical to examine variables that may contribute to successful implementation of high-quality reading interventions in the context of RtI.

Understanding the barriers and facilitators to implementation are critically important considering that the success of the RtI movement hinges on the uptake of empirically-based interventions in the context of routine practice. Many variables are thought to relate to successful implementation, such as: time, resources, leadership, planning, preparation of professionals, empirical evidence, and evaluation (notably treatment integrity) (Jimerson, Burns, & VanDerHeyden, 2007; Kovaleski, 2007; Fuchs, Fuchs, & Compton, 2012). These variables are critical factors thought to enable the utilization of quality reading interventions; however, it remains unclear how these change mechanisms may contribute to the ability to implement specific quality components of reading interventions. Therefore, it is necessary to empirically examine the potential facilitators and barriers of high-quality reading interventions in the context of RtI. Specifically, the following four components are indicators of quality reading intervention implementation in the context of RtI: 1) the use of an evidence-based reading intervention, 2) the use of psychometrically sound progress monitoring, 3) the use of treatment integrity measures, and 4) the use of a trained interventionist.

Evidence-Based Reading Interventions

Evidence-based practice begins with the use of the best available research evidence (i.e., evidence-based interventions) (APA Presidential Task Force on Evidence-Based Practice, 2006). Evidence-based interventions are distinguished by being based on research findings that demonstrate they are likely to produce predictable, beneficial, and effective results (Forman & Burke, 2008). A considerable body of scientific evidence drawn from a variety of research designs and methods attests to the availability of effective practices.

The available literature suggests that targeted reading interventions benefit students at risk for reading difficulties. Tier 2 interventions have been shown to remediate reading difficulties and prevent later reading problems (Chard et al., 2008; Goss & Brown-Chidsey, 2012; VanDerHeyden et al., 2007; Wanzek & Vaughn, 2008). It is recommended that instruction is provided by a school professional with specialized training who works with small homogeneous groups (4-6 students), 4 days per week, and approximately 30 minutes per session (Fuchs, Fuchs, and Stecker, 2010; Gross & Brown-Chidsey, 2012). While less empirically studied, for those students who are non-respondent to an initial intervention, there is evidence that when teachers provide intervention at greater intensity, reading outcomes can be improved (Al Otaiba et al., 2011; Conner, Morrison, Fishman, Schatschneider, & Underwood, 2007; Denton et al., 2013). The intensity of interventions can be increased at Tier 3 by decreasing group size and increasing time in intervention (Denton et al., 2013; Vaughn, Denton, Fletcher, 2010).

This research establishes a foundation for the provision of evidence-based reading interventions, however, best research evidence may include many different research designs, including: clinical observation, qualitative designs, systematic case studies, single-case experimental designs, ethnographic research, process-outcome studies, interventions in naturalistic settings, randomized controlled trials (RCTs), and meta-analyses, with some methods better suited to address certain types of questions (Stoiber & Desmet, 2010). Despite general evidence of effectiveness, it is critical to realize that evidence and research will not address all practice needs and must balance internal and external validity. For example, in many cases it is unclear the degree to which results of efficacy and effectiveness research can be generalized from primarily White samples to minority and marginalized populations. There are many other considerations in the process of determining how to integrate research into day-to-day practice, such as: a) the relative weight to place on different research methods; b) the representativeness of research samples; c)whether research results should guide practice at the level of principles of change, intervention strategies, or specific protocols; d) the generalizability and transportability of treatments supported in controlled research to clinical practice settings; and e) the extent to which judgments can be made about treatments of choice when the number and duration of treatments tested has been limited (APA Presidential Task Force on Evidence-Based Practice, 2006).

In light of these complexities and the growing evidence-based movement, it is not surprising that various professional groups within the field of psychology and education (including school psychology) have developed criteria and published information regarding evidence-based interventions (Stoiber & Desmet, 2010; Coffee, Newell, &

Kennedy, 2014). The American Psychological Association has approved and posted criteria for evaluating treatment standards, and a manual has been developed for school psychology to identify, review, and code studies of psychological and educational interventions for various academic, social-emotional, and mental health concerns experienced by school-age children and their families (titled: *Procedural and Coding Manual for the Review of Evidence-based Interventions,* Kratochwill & Stoiber, 2002). In addition to coding manuals or guidelines, multiple groups have begun reviewing academic and behavioral programs to support implementation in the schools by providing evidence ratings on the level of research to support practices on a variety of topics (Slavin, 2008).

The purpose of these groups is to educate professionals on the criteria to be considered for the evaluation of interventions. In particular, the evaluation of specific interventions should promote the implementation of effective practices by focusing on interventions utilizing reliable and valid methods that led to successful intervention outcomes (Slavin, 2008). The use of reliable and valid methods should help standardize approaches and reduce practice variation. However, different groups utilize different criteria to evaluate intervention effectiveness and also have different criteria in place to determine whether a study has met the requirements for review (Stoiber & Desmet, 2010). While many groups focus on mental health, the following groups review reading interventions in an ongoing fashion: What Works Clearinghouse, Promising Practices Network, and Johns Hopkins Best Evidence Encyclopedia. Other groups, such as the National Reading Panel, Texas Reading First Initiative, Florida Center for Reading Research, and Oregon Reading First Center have created reading reports on effective

reading interventions, but they are not reviewed here since they represent one-time 'snapshot' reviews of reading intervention evidence (i.e., they are not reviewing interventions in a continuous manner) (Kratochwill, Clements, & Kalymon, 2007).

What Works Clearinghouse. The What Works Clearinghouse (WWC), part of the U.S. Department of Education's Institute of Education Sciences (IES), focuses on reviewing and disseminating the evidence of effectiveness on student-related outcomes in order to enable educators to access and reach sound conclusions about the effectiveness of interventions (What Works Clearinghouse, 2010). The WWC provides a handbook to document their procedures and criteria for review. After selecting a topic and systematically reviewing / screening the literature, WWC reviewers examine studies for which there is high or moderate confidence that the outcomes can be attributed solely to the intervention rather than to other factors (i.e., confounds). Accordingly, randomized controlled trials and quasi-experimental designs with equivalent groups are the only type of studies that are considered (What Works Clearinghouse, 2010).

Randomized trials with low attrition provide strong evidence (meets evidence standards), while randomized trials with high attrition provide weaker evidence (meets evidence standards with reservations). The impact of attrition is described in greater detail in the WWC manual, but it is based on overall attrition rate and differential attrition rates (i.e., extent to which the attrition may relate to the student outcomes in the topic area) (What Works Clearinghouse, 2010). For non-randomized trials, studies with equivalent groups are considered weaker evidence (meets evidence standards with reservations), while non-randomized trials without equivalence are not considered for

review (does not meet evidence standards). Equivalence must be demonstrated on observed characteristics defined in the topic area (What Works Clearinghouse, 2010).

Beyond these standards, a study may not meet standards for many other reasons, such as: 1) it does not include a valid or reliable outcome measure, 2) the intervention and comparison groups are not shown to be equivalent, 3) the results may be due to potential confounds, 4) it includes only outcomes that are over-aligned with the intervention, 5) estimates of effects did not account for differences in pre-intervention characteristics, 6) there was only one unit of analysis in one or both conditions, 7) the intervention was combined with another intervention, or 8) the intervention was not implemented as designed (What Works Clearinghouse, 2010).

If an intervention has at least one study meeting standards or meeting standards with reservations, an intervention report is prepared that presents the empirical findings, the rating of the evidence, and the improvement index, which represents the magnitude of the effect synthesized from the evidence. Effect sizes are typically based on the standardized mean difference except when a finding is based on a misaligned analysis or when a study examines multiple outcomes simultaneously (see report for details). Overall, ratings are provided based on the following criteria. 'Positive effects' requires two or more studies showing statistically significant positive effects (at least one with a strong design) and no studies showing statistically significant or substantively important negative effects. 'Potentially positive effects' must include at least one study showing a statistically significant or substantively important positive effect and no studies showing statistically significant or substantively important negative effects. 'Mixed effects' are when at least one study shows statistically significant or substantively important positive

effects and at least one study shows statistically significant or substantively important negative effects (but cannot be more than the number of positive results); mixed effects may also occur if one study shows statistically significant or substantively important effects and more studies show an indeterminate effect than a statistically significant or substantively important effect. 'No discernible effects' is when none of the studies show a statistically significant or substantively important effect. The WWC also offers 'potentially negative effects' and 'negative effects', which are the same criteria as for potentially positive and positive but in the reverse direction (What Works Clearinghouse, 2010).

Promising Practices Network. The Promising Practices Network requires the use of randomized-control or quasi-experimental designs; however, it does not require the verification of group equivalence on pre-existing variables. Also, even if the specific goal of the program does not address an indicator, but a positive effect is demonstrated, it will be included under the indicator of effectiveness. In this manner, the Promising Practices Network is more lenient than the What Works Clearinghouse in that studies with some weaknesses are still reviewed (Promising Practices Network, 2014).

The Promising Practices Network denotes three types of evidence levels. A 'proven program' is one with a substantial effect size (at least one outcome changes by 20% or .25 standard deviations) and is statistically significant at the .05 level; it must also utilize a convincing comparison group to identify program impacts (which may include quasi-experimental designs) and have a sample size exceeding 30 in both the treatment and comparison groups. A 'promising program' is designated for a program that has a positive effect size (change in outcome is more than 1%) and is significant at the .10

level. Further, the intervention may impact an intermediary outcome (for which there is evidence that it is associated with one of the promising practice network indicators), the study has a comparison group (but it may exhibit some weaknesses), and the sample size of evaluation exceeds 10 in both the treatment and comparison groups. If a program does not meet these conditions, it will not be listed on their website (Promising Practices Network, 2014).

Johns Hopkins Best Evidence Encyclopedia. The Johns Hopkins Best Evidence Encyclopedia reviews meta-analyses or other quantitative syntheses that apply consistent, scientific standards to bodies of evidence that are of high methodological quality and evaluate realistic implementations of programs currently available to educators. More specifically they must: 1) consider all studies in the area and conduct an exhaustive search for studies meeting well-justified standards of methodological quality, 2) present quantitative summaries of evidence on the effectiveness of programs with children in grades K-12, focusing on achievement outcomes, 3) focus on studies comparing programs to control groups, with random assignment to conditions or matching on pretests indicating equivalency prior to treatment, 4) summarize programs in terms of effect sizes and statistical significance, 5) focus on studies that took place over periods of at least 12 weeks (to avoid brief artificial laboratory studies), and 6) focus on studies that used measures that assessed the content studied by control as well as experimental studies in order to avoid studies that used measures inherent to the experimental treatment (Best Evidence Encyclopedia).

Reviewers from the Best Evidence Encyclopedia then rate educational programs according to the overall strength of the evidence supporting their effects on student

achievement. 'Strong evidence' of effectiveness is designated for programs with at least one large randomized (or randomized quasi-experimental) and one additional large qualifying study (or multiple smaller studies), with a combined sample size of 500 and an overall weighted mean effect size of at least .20. 'Moderate evidence' of effectiveness is designated for programs with two large matched studies (or multiple smaller studies) with a collective sample size of 500 and with a weighted mean effect size of at least .20. 'Limited evidence: strong evidence of modest effects' is the same as the criteria for moderate evidence except that the weighted mean effect size ranges from .10 to .19. Lastly 'limited evidence: weak evidence with notable effects' is designated for programs with an effect size of at least .20, but have an insufficient sample size to meet criteria for moderate evidence of effectiveness (Best Evidence Encyclopedia).

Summary. One challenge to the simple adoption of the WWC criteria is that different reviewing groups typically adopt their own set of review criteria. With these differing criteria, the groups may come to different conclusions (e.g., one organization's 'promising program' is another group's 'effective program'). So, while WWC represents a large, robust, and rigorous process for the evaluation of interventions, it is necessary to consider other review groups. The What Works Clearinghouse, Promising Practices Network, and the Johns Hopkins Best Evidence Encyclopedia represent a summary of evidence-based reading interventions because they provide three different robust methods of review (mix of criteria for research controlled trials, quasi-experimental designs, and meta-analyses) and also provide descriptive summaries of evidence. Collectively, they represent the best available research evidence (i.e. evidence-based interventions), and therefore their use should promote the implementation of effective practices as these

interventions have been shown to utilize reliable and valid methods that led to successful intervention outcomes in reading. While the use of an evidence-based intervention is critical, progress monitoring is necessary for evaluating and improving program effects (Deno, 2003; Deno, 2005). It allows the determination of whether the intervention is successful in its particular context, both formatively and summatively (Upah, 2008).

Progress Monitoring

Progress monitoring is the process of assessing students' academic performance on a regular basis (Fuchs & Fuchs, 2002). In the context of evidence-based intervention implementation, it is critical because it allows the documentation of student growth and a determination of whether children are appropriately benefiting from the given instructional program (i.e., evidence-based intervention). Progress monitoring enables the assessment of student performance over time, the quantification of rates of improvement, and evaluation of instructional effectiveness (National Center on Response to Intervention, 2010). Specifically designed progress monitoring tools are preferred for the aforementioned purposes because more traditional school-based assessments (e.g., state standardized tests) and measures designed to be used solely for screening purposes lack sensitivity to short-term change, and do not have multiple equivalent forms for monitoring progress (National Center on Response to Intervention, 2010).

Progress monitoring in reading is commonly described in two forms: 1) mastery measurement and 2) curriculum-based measurement (CBM) (Fuchs & Fuchs, 2002). Curriculum-based measurement is a set of standardized and short duration tests that can be used by educators to evaluate the effects of instructional interventions (Shinn, 2008). However, most classroom assessment relies on mastery measurement, which is when

students are tested on their mastery of a single skill, and upon demonstrating sufficient levels of this skill, they move onto the next skill. At different times of the school year, different skills are being assessed (Fuchs & Fuchs, 2002). Of course, due to the changing nature of the skill, the difficulty of the tests keep changing with successive mastery, rendering test scores from different times of the school year incomparable, and it becomes impossible to quantify or describe rates of progress. Further, mastery measurement has unknown psychometric properties (i.e. reliability and validity) and they fail to provide information about whether students are maintaining previous skills (Fuchs & Fuchs, 2002). These criticisms, along with strong empirical support, provide the rationale for the use of curriculum-based measures for progress monitoring (Fuchs & Fuchs, 2002).

Psychometric properties. Psychometric properties refer to the traits or characteristics of a particular measurement tool, which is commonly defined by reliability and validity. Curriculum-based measures are standardized (with established norms and criterion) and have well-documented psychometric properties (Fuchs & Fuchs, 2002). Because they were originally designed as a method for monitoring student growth and evaluating the effects of instructional programs, it was important (in their early development) to ensure that the measures have documented reliability and validity (Deno, 1985). In 1989, Marston (as cited in Wayman et al. 2007) reviewed the reliability and validity of CBM in the context of reading, which took the form of two measures, word identification and reading aloud (oral reading fluency). The results of five studies demonstrated strong test-retest reliability, with coefficients ranging from .82 to .97 (most coefficients above .90); alternate-from reliability coefficients ranged from .84 to .96 with

most coefficients above .90 as well; lastly inter-rater agreement was .99. Fourteen additional studies from Marston (as cited in Wayman et al. 2007) examined the validity of the curriculum-based measures. They found that CBM correlated well with other published measures of reading (coefficients ranged from .63 to .90; most above .80) and basal reading mastery measures (coefficients ranged from .57 to .86, with half above .80). Further, Marston (1989) demonstrated that reading aloud correlated with teacher judgment and various measures of reading comprehension, discriminated between low and high performing students, and was sensitive to growth. The Marston (1989) review provided preliminary support for the technical adequacy of the measures; however, the research on CBM has expanded considerably since this time necessitating a more recent review.

In 2007, Wayman et al. reviewed all empirical studies of CBM that focused on the validity of reading measures that were published since the time of Marston's (1989) review. They focused on studies of school-age students and the technical adequacy of the most commonly used reading measures: reading aloud, maze selection, and word identification. Reading aloud, also known as oral reading fluency (ORF) has students read aloud from a passage, typically for 1 minute, and the number of words read correctly is scored. Omissions, insertions, substitutions, hesitations, and mispronunciations are marked as errors. In maze, students read through a passage in which every seventh word has been deleted and replaced with three word choices (one correct choice and two distracters). Students read the passage silently, typically for 1 to 3 minutes, and make selections as they read. The outcome measure is the number of correct selections. Lastly, in word identification, students read aloud from a list of high frequency words

(usually 1 minute), and the number of words read correctly is scored. Similar to reading aloud, omissions, insertions, substitutions, hesitations, and mispronunciations are marked as errors (Wayman et al., 2007). The following subsections are more detailed explanations of the Wayman et al. (2007) findings.

Reading aloud (oral reading fluency). Despite the preliminary support for the technical adequacy of oral reading fluency, practitioners and researchers continued to express doubts about the relationship between reading aloud for 1 minute and reading proficiency, particularly in terms of reading comprehension. Multiple approaches were devised to tease apart the relationship between reading aloud and reading comprehension (Wayman et al., 2007). For example, Fuchs, Fuchs, and Maxwell (1988) compared the validity of reading aloud to other measures of comprehension (e.g. story retell, and question-answering measures) and found that reading aloud scores actually correlated more strongly with scores on comprehension and word skills subtests of a standardized achievement test than these 'typical comprehension measures'. Kranzler, Brownell, and Miller (1998) took a different approach examining the idea that the number of words read aloud in 1 minute might be a reflection of processing speed and not necessarily reading comprehension. Multiple regression analyses revealed a significant relationship between reading aloud and reading comprehension measures. Specifically they found that 11% of the variance was uniquely explained, and could not be explained by general cognitive ability or speed and efficiency of cognitive processing. These results suggest that reading aloud is more than just a measure of fluent decoding and also more than just a measure of general cognitive processing speed (Wayman et al., 2007).

Maze. Despite the relatively strong support for reading aloud, it has been shown to be less appropriate for older students. Accordingly, the maze measure has been examined as a 'purer' reading comprehension measure, and one that is more appropriate for older students. Espin, Deno, Maruyama, and Cohen (1989) reported on the preliminary technical adequacy of the measure. On a sample of 2,000 students, correlations between maze and reading aloud for grades 3, 4, and 5 ranged from .77 to .86 and data from the entire sample revealed a stable pattern of increase in maze scores from grades 1 to 6 as well as from Winter to Spring within each grade. Extending this research, Fuchs and Fuchs (1992) compared maze to other alternative measures (question answering, story recall, and cloze). The study revealed that the maze task was sensitive to change in performance over time, and, relative to other measures, had a relatively small ratio of slope to standard error of the estimate, thereby making it easier to detect growth (Wayman et al., 2007).

Word identification. While reading aloud has emerged as a more commonly used measure, there was interest in word identification at younger age levels because many first grade students would be unable to read any words from text in the fall, creating a floor effect (Wayman et al., 2007). Words presented in a word identification task can be controlled for difficulty and do not require a coherent story (Wayman et al., 2007). In a comparison of multiple early reading measures, including word identification, letter reading, letter copying, letter-sound production, and letter-sound selection, word identification and letter reading produced the most technically adequate data (Daly, Wright, Kelly, & Martens, 1997). Word identification specifically demonstrated strong test-retest reliability (.94), and provided concurrent validity through its correlation with

broad reading on the WJ-R (correlation of .40) and predictive validity through its correlation with passage reading and word identification 4 months later (correlation of .73). More recently, Compton, Fuchs, Fuchs, and Bryant (2006) examined the use of a word identification measure in the context of a response-to-intervention approach for first-grade students. Students were administered a prediction battery in fall of first grade and also progress-monitored for a 5-week period using the word identification task to calculate a slope of improvement. These students were followed until the end of second grade to examine the utility of word identification level, slope and a combination of the two as part of a process for predicting performance at the end of second grade. The results demonstrated that adding word identification level and slope significantly improved classification accuracy for the identification of at-risk students beyond the use of phonemic awareness, rapid naming, and oral vocabulary measures (Compton, Fuchs, Fuchs, & Bryant, 2006).

Predictive utility. In addition to their psychometric properties, curriculum-based measures assess student progress toward long-term and more general goals, rather than mastery of successive objectives. Multiple alternate forms of these short tests are developed that sample performance toward the long-term goal, not just the content or skills that the student is currently learning (Fuchs & Fuchs, 2002). For example, teachers can use story passages outside the student's curriculum as an indicator of overall improvement in reading as the general outcome, as opposed to just successful reading of particular passages within a student's curriculum. The relative advantage is that performance on CBM illustrates what a student is able to do relative to that long-term goal. Because the content or level of difficulty on the measures and time allotted for the

assessment tasks remain constant, CBMs offer a method of assessment providing data on student performance over time. Thus, CBM provides an ideal method of measuring growth to predict whether students are on target towards the long term goal (Fuchs & Fuchs, 2002).

Early studies of predictive utility focused on the predictive capacity of CBM in for state reading tests. They focused on establishing benchmark scores that would predict passing or failing a state reading test. Since testing for statewide accountability tests begins at 3rd grade, oral reading fluency has been the measure most commonly utilized for these predictive purposes as it is the strongest CBM measure from 1st through 4th grade. Good, Simmons, & Kame'enui (2001) utilized a series of linked, short-term, longitudinal studies of four cohorts to examine the strength of relations and performance probabilities among oral reading fluency measures and a third-grade high-stakes reading assessment, which demonstrated particularly strong linkages to subsequent performance. They found a statistically significant relationship between students who reached benchmark during the Spring of 1st grade and their performance on the benchmark during the Spring of 2nd grade. Specifically, the score in Spring of 1st grade explained 67% of the variance in the score at Spring of 2nd grade (97% of students reaching 1st grade benchmark met 2nd grade benchmark). The strong relationship between 1st and 2nd grade scores is of greater significance since the oral reading fluency (ORF) benchmark can be used to predict passage of the third-grade high-stakes reading assessment. Specifically, ORF in the Spring of 3rd grade explained 45% of the variance on the state standardized test, with 96% of students who attained benchmark at Spring of 3rd grade achieving proficiency on the state standardized test. Collectively, this provides emerging support

for the predictive utility of ORF for later reading proficiency on ORF measures, which in turn is predictive of performance on state standardized tests. Other studies have demonstrated similarly strong correlations that link performance on CBM measures to scores on state standardized tests (Crawford, Tindal, & Stieber, 2001; Hintze & Silberglitt, 2005).

Later studies on the predictive capacity of CBM (ORF specifically) have focused on sensitivity, specificity, positive predictive power, and negative predictive power (Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Silberglitt & Hintze, 2005). Sensitivity is the percentage of students below a cut score who fail a test, while specificity is the percent of students above a cut score who pass a test. Positive predictive power is the probability that a student with a score below the cut score will truly fail the test, and negative predictive power represents the probability that a student with a score above the cut score will truly pass the test. In their review, Wayman et al. 2007 found that correlations between scores on oral reading fluency measures and various state reading tests generally ranged from .60 to .80 (from four studies). Further, the diagnostic efficiency statistics were fairly consistent across these studies: sensitivity ranged from 65% to 76%, specificity ranged from 74% to 82%, positive predictive power ranged from 55% to 77% (with one exception at 41%) and negative predictive power ranged from 83% to 90% (with one exception at 46%). Collectively, these studies demonstrated the promising predictive utility of oral reading fluency.

In sum, oral reading fluency has shown to be predictive of state standardized tests, particularly from 1st to 3rd grade (when state standardized tests begin). Multiple studies have demonstrated subsequent relationships between and within grades, which in turn

correlate to state test results. Further studies have demonstrated that oral reading fluency can correctly identify students who are below a cut score and then fail the state test (sensitivity), and students above a score who then go on to pass the state test (specificity).

Decision-making with CBM data. Curriculum-based measures were originally developed to inform instructional decision-making in special education (Deno, 1985), which at least in part, explains their prominent use in response-to-intervention systems. RtI models are similar in that decision-rules are applied to the students' response to empirically-based interventions. Students are considered for special eligibility if they continue to show a lack of adequate response to the targeted evidence-based interventions (Deno, 2005). In RtI, one key distinction from its original purpose is that since students may be considered for special education based on a lack of response, it is of heightened significance that the process of using growth to inform decision-making is a valid and reliable process.

A review by Ardoin, Christ, Morena, Cormier, & Klingbeil (2013), sought to provide a review of CBM literature to make recommendations regarding decision rules for progress monitoring, and to review the underlying evidence supporting those recommendations. Specifically, they examined the validity of decision rules and evidence that serves as the basis for those decision rules. Focusing on CBM-R, Ardoin et al. (2013) searched ERIC, Psycinfo, and Academic Search Complete to identify any journal article, instructional manual, and chapter related to CBM-R. After applying exclusionary criteria (e.g., subjects other than reading and conference presentations), 102 documents were reviewed for coding. They examined the document type (e.g. chapter or manual), decision rules discussed (data point and trend line decision rules), recommended

number of CBM-R data points, accuracy of CBM-R decision rules, and accuracy of CBM-R growth estimates. Data point rules involved the recommended number of data points to be collected for decision-making purposes while trend line recommendations involved the method for drawing the trend line.

Of the 102 documents reviewed, 78 discussed the use of CBM-R data for progress monitoring data and 60 of those documents explicitly stated how to employ one or more CBM-R decision rules. Specifically, 7 discussed the use of data point decision rules, 47 discussed trend line decision rules, and 22 discussed both. Of the articles discussing data point rules recommendations ranged from 3 to 20, with 7 data points being the most frequent. Of the articles discussing trend lines, the majority (41) recommended the use of ordinary least squares for drawing the trend line. Despite the numerous recommendations in applying decision rules to CBM-R progress monitoring, the review did not identify a single study that evaluated the accuracy of the data point or trend line decision rules. Accordingly, the use of CBM progress monitoring as a decision-making tool in the evaluation of instructional effects is not a strongly evidence-based practice (Ardoin et al., 2013).

That being said, the use of progress monitoring for this purpose is a theoretical foundation of the model, and there is emerging evidence to support its use. A sequence of five studies by Christ et al., (2013) sought to more clearly specify progress monitoring procedures that would produce reliable and valid slopes for use with instructional decision-making. In light of the criticism highlighted by Ardoin et al. (2013), the Christ et al. (2013) sequence of studies examined many components of progress monitoring procedures in order to determine the conditions that promote valid and reliable slopes for

instructional decision-making purposes. Specifically, they examined how the schedule (range = 1 per day to 1 per month), duration (range = 2 to 20 weeks), and number of CBM per occasion (1 or 3), my impact the quality of slope estimates. They defined the quality of the slope estimates based on the reliability, validity, and diagnostic accuracy (sensitivity and specificity) of the estimated slopes. The five studies were examinations of different combinations of CBM-R schedule, duration, and number per occasion, and their impact on the dependent variables.

A linear mixed effects regression model was used to simulate true and observed CBM-R progress monitoring data based on a large dataset of second (n = 1517) and third grade students (n = 1561). The five studies examined the number of weeks to reach sufficient reliability/validity and diagnostic accuracy for the following administration schedules: 1) one CBM-R per occasion 5 times per week, 2) one CBM-R per occasion 3 times per week, 3) three CBM-R per occasion 2 times per week, 4) three CBM-R per occasion once per week, and 5) three CBM-R per occasion once per month. They found that conditions 1 and 3 led to reliable, valid, and diagnostically accurate slopes in the shortest amount of weeks. Specifically, for both studies (1 & 3) reliability and validity for low-stakes decisions was established after 8 weeks and after 12 weeks for high-stakes decisions, while diagnostic accuracy was established for low-stakes decisions in 5 to 6 weeks and after 8 weeks for high-stakes decisions. The other conditions required 2 to 5 more weeks of data collection to reach slopes of sufficient quality. It is important to note when interpreting these findings that these occurred with very good data quality (near ideal data collection conditions); with good quality data (slightly less than ideal and likely more similar to typical contexts), it typically took 4 (or more) weeks longer to

reach slopes with sufficient quality. The key finding from these simulations was that progress monitoring should occur much more frequently than is commonly recommended. Specifically, their results reiterated the importance of progress monitoring under best conditions (high-quality probe sets and tightly controlled conditions) and also suggested that multiple data points be collected per week to inform instructional decision-making (Christ et al., 2013). Further, the results indicate that the frequency of progress monitoring is a critical component of quality, since the frequency impacts the reliability and validity of the slopes of progress used for decision-making purposes (Christ et al., 2013).

These findings are generally consistent with a review of the basic components of response-to-intervention by the Institute for Education Sciences (2009). They adhered to similar guidelines as the WWC Evidence Standards to assess the quality of evidence supporting the following components of RtI: universal screening, differentiated reading instruction for all students, small-group interventions (i.e., Tier 2), progress monitoring at least once a month, and intensive interventions (i.e., Tier 3). Aligned with WWC standards, they reviewed randomized controlled trials and quasi-experimental design studies. Of the 11 studies that evaluated effects of Tier 2 interventions, only 3 reported using mastery checks or progress monitoring in instructional decision-making, with none of them demonstrating that progress monitoring is essential. The review does highlight that there is theoretical support for progress monitoring and cited that one study demonstrated that progress monitoring (in oral reading fluency or word identification) increases teachers' awareness of students' current level of reading proficiency and has a positive effect on the instructional decisions teachers make. Overall, they found that

there is a low level of support for this practice, which means that it has yet to garner sufficient empirical support but is based on expert opinion and theory. Accordingly, it remains a recommended practice and the more recent work by Christ et al. (2013) highlights that support for this practice is emerging, so long as the progress monitoring occurs frequently under tightly controlled conditions.

Summary. Using curriculum-based measures for progress monitoring has multiple validated advantages relative to mastery measures in the context of intervention implementation. Curriculum based measures have documented psychometric properties, have shown predictive utility, and coupled with instructional modification, can positively impact student performance (Stecker, Fuchs, & Fuchs, 2005; Wayman et al., 2007). Further, there is theoretical support for the use of CBM as a progress monitoring tool in the context of instructional decision-making and emerging empirical evidence supporting its use as well (Christ et all., 2013). Specifically, to ensure interpretable slopes of progress (i.e. reliability and validity of slopes), progress monitoring should occur at least two times per week for low-stakes decisions (such as instructional modification in Tier 2), and it should occur more frequently for high-stakes decisions, such as the determination of special education eligibility. Accordingly, at the individual level (Tier 3) progress monitoring should occur five times per week. Progress monitoring allows the determination of whether an intervention is successful in its particular context (Fuchs & Fuchs, 2002), however, in order to attribute the effects to the intervention, it is necessary to ensure the integrity of the intervention (Hagermoser Sanetti & Kratochwill, 2009; Noell & Gansle, 2006).

Treatment Integrity

Historically, treatment integrity has been assumed rather than assessed in research and practice across many fields, but more recently there has developed an emerging consensus in education regarding the need to address treatment integrity (Hagermoser Sanetti & Kratochwill, 2009). Also referred to as treatment fidelity, intervention integrity, and procedural reliability, treatment integrity is defined as the degree to which an intervention is implemented as planned (Hagermoser Sanetti & Kratochwill, 2009). The role of treatment integrity has garnered considerable support over the past 5 years, as it is related to the discrete delivery of evidence-based interventions. Specifically, if critical components of evidence-based interventions are not implemented, the intended recipients are unlikely to benefit (Hagermoser Sanetti & Kratochwill, 2014; Noell & Gansle, 2006). Further, among the many variables that have been identified as influential in the efficacy of interventions, treatment integrity is the most proximal and definitive (Hagermoser Sanetti & Kratochwill, 2014).

As an emergent topic of research, it is not surprising that reviews of treatment literature over the past 15 years have indicated minimal consideration of treatment integrity (on average, 19.9% of treatment outcome articles included treatment integrity data) (Hagermoser Sanetti & Kratochwill, 2008). While this finding applies to empirical research, a recent survey of nationally certified school psychologists indicated that a similar situation is apparent in school-based practice. Respondents reported that treatment integrity is important to consider when evaluating intervention effectiveness (56.2%), and yet, as part of their consultation with teachers, only 11.3% of respondents always assessed treatment integrity and 41.6% sometimes assessed treatment integrity (Cochrane

& Laux, 2008). The results are even bleaker in the context of a school-based problem-solving team process, with 1.9% always and 40.4% sometimes assessing treatment integrity (Cochrane & Laux, 2008). It is readily apparent from this research that treatment integrity is underutilized in both research and practice contexts, however, it remains necessary to verify empirically that for an intervention to be effective it must be implemented as intended.

Integrity in school contexts. In school-based contexts, the measurement of treatment integrity is significant because it enables the interpretation of program/intervention effectiveness (O'Donnel, 2008). Integrity can reveal important information in terms of confidence that the observed outcomes can be attributed to the intervention. Without assurances of integrity, it may be unclear how to account for negative or ambiguous findings because they may be due to an ineffective program, or due to the failure to implement the program as intended. Integrity in an effectiveness study helps promote external validity and has implications for scaling up the practice (O'Donnel, 2008).

O'Donnell (2008) conducted a review of how integrity of implementation is defined, conceptualized, measured and applied in the context of K-12 core curriculum interventions and outcomes. Articles were included if they defined and conceptualized fidelity of implementation within efficacy and effectiveness studies. In their search, the literature on integrity of implementation was not very large, with the majority of the articles pertaining to public and mental health fields. Of an initial search yielding 133 studies (ProQuest and EBSCOhost), 23 pertained to curricular interventions (a criteria/focus of the study). After review, only 5 studies were conducted in K-12 settings

and examined the efficacy or effectiveness of core curriculum interventions that could be adopted and implemented by a single teacher in a classroom. The key finding was that all five studies consistently showed significantly higher outcomes when the program was implemented with greater integrity (O'Donnel, 2008).

This evidence provides support for the idea that higher levels of treatment integrity result in better outcomes, which has also been verified in a broader review that examined the effects of integrity on prevention and health promotion programs for children and adolescents on the following topics: physical health and development, academic performance, drug use, and various social and mental health issues such as violence, bullying, and positive youth development (Durlak & Dupre, 2008). Durlak & DuPre (2008) searched Psycinfo, MEDLINE, and Dissertation abstracts to find prevention and health promotion programs for the aforementioned topics and found nearly 500 relevant interventions (from 5 meta-analyses) and 59 additional quantitative studies. They found that mean effect sizes were two to three times higher when programs were carefully implemented and free from serious implementation problems, while the 59 additional studies confirmed higher levels of implementation were often associated with better outcomes, particularly when fidelity or dosage was assessed. The review provides similarly supportive results that higher levels of treatment integrity equates to more positive student outcomes (Durlak & Dupre, 2008).

Lastly, a review of intervention studies that target school-related problems was conducted by Noell and Gansle (2006). The focus of the study was on the degree of treatment plan implementation needed for successful outcomes, with studies examining disruptive behavior, anxiety, peer tutoring, social skills training, differential

reinforcement, strategy instruction, and multi-systemic therapy for juvenile offenders (Noell & Gansle, 2006). The studies of interest either manipulated the integrity of the intervention or simply measured the integrity of the intervention. Across the studies reviewed, lower treatment integrity was associated with lower effect sizes of the treatment (Noell & Gansle, 2006). Taken collectively, the three studies described above provide strong empirical support for the relationship between treatment integrity and treatment outcomes; however, a challenge to integrity assessment is that methods of measurement are far from standardized (Hagermoser Sanetti & Kratochwill, 2014).

Measuring integrity. Despite support for the importance of treatment integrity, it is not clear how best to operationalize the components of an intervention for the purposes of integrity assessment, and there is an abundance of methods available including: systematic observation, videotaped observation, rating scales completed by the interventionist or observer, behavioral interviews with the interventionist, and permanent products (Hagermoser Sanetti & Kratochwill, 2014). The varying methods of integrity measurement are typically conceptualized as direct or indirect, with each approach having its own advantages and disadvantages (Hagermoser Sanetti & Kratochwill, 2014).

The direct assessment of treatment integrity is based on the observation of treatment implementation as it is being implemented (e.g. systematic observation, and videotaped observation, which may also include rating scales completed by the observer). It is similar to the systematic observation of behavior in applied settings, with several factors necessary for consideration, such as: purpose of the observations, content of the observations, amount of behavior to be observed, and the quality of the data produced. The primary disadvantages of direct approaches are the potential reactive effects of the

observer's presence (e.g. social desirability), and the considerable resources (i.e. time) required to have an independent observer assess intervention integrity. Unfortunately, little empirical research has examined reactivity in the context of an integrity assessment, and it is also unclear how much observation data may be necessary to produce a representative sample of treatment integrity (Hagermoser Sanetti & Kratochwill, 2014).

Indirect assessment of treatment integrity is based on methods of assessment that occur subsequent to its implementation (e.g. self-reports, behavioral interviews, and permanent products). These methods have the general advantage of being less time-consuming, but are also subject to some of the same drawbacks as the direct assessment methods. For example, self-report and interview measures may also produce reactive effects such that the implementer may report high levels of integrity and yet implement with low levels. Some treatments leave a permanent product in the environment that can be used to assess integrity. However, these methods typically only provide a general estimate of integrity since each component may not have a permanent product (Hagermoser Sanetti & Kratochwill, 2014).

Unfortunately, there is little research on how to best utilize the multiple available integrity assessment methods. Despite this, it is known that multiple methods can be effectively used to assess treatment integrity (Hagermoser Sanetti & Kratochwill, 2014). Progress monitoring and treatment integrity enable the determination of whether an intervention is successful, but an additional contributor to the successfulness of the intervention (beyond its empirical support) involves the training of the interventionist.

Trained Interventionist

Similar to treatment integrity, the use of a trained interventionist has intuitive appeal in terms of its relationship to student outcomes; it would be expected that someone with specific training in the implementation of a particular intervention would be more effective than someone without training. Studies of preventive interventions comparing teachers who received in-service training to those without training demonstrate that training, in general, is an important element for effective implementation (Perry et al., 1990; Ross et al., 1991). For example, Ross et al. (1991) conducted an experimental sub-study of the *Teenage Health Teaching Modules* curriculum. The curriculum was designed to require no teacher training, but they examined whether training enhanced teacher preparedness to teach the curriculum and whether teachers who were trained implemented the curriculum more successfully. Eighty-five teachers were randomly assigned to a training (3-day seminar) or no training condition. Results demonstrated that training led to higher self-reported preparedness and had positive effects on curriculum implementation (percentage of required activities taught; $p < .005$), student outcomes (based on knowledge scores; $p < .001$), and fidelity of implementation (lack of modification to curriculum; $p < .01$) (Ross et al., 1991). In addition to promoting greater success in implementation, the use of a trained interventionist has been shown to impact the transportability and sustainability of the intervention and to promote integrity.

Training for transportability and sustainability. For many years, it was thought that if a practice or program was efficacious and if information were made available about it, it would be implemented. Despite the availability of and policy support for evidence-based interventions, a significant body of literature indicates that

implementation of evidence-based interventions in schools is low (Forman et al., 2009). This challenge refers to the transportability and sustainability of interventions, such that interventions must be successfully 'transported' from research settings to practice settings, and 'sustained' beyond their original adoption. In order for evidence-based interventions to be transported and sustained in the practice context interventionist training is a necessary component (Fixsen, 2005).

A review by Forman et al. (2009) examined the factors that are important to successful implementation and sustainability of evidence-based interventions in school settings. They found 98 interventions that were school-based, tested in studies using either randomized control or quasi-experimental designs, and provided outcome data showing clear evidence of program's effectiveness. From these 98 interventions, 29 were endorsed by four or more programs (i.e., programs that endorsed interventions as 'evidence-based,' such as the Promising Practices Network, collaborative for academic, social, and emotional learning, among others), which formed that basis of the analysis. Each of the intervention developers of these programs were approached for interview (25 agreed, 86% response rate) to discuss barriers and facilitators to implementation of the intervention and sustainability. Results indicated that good training was an essential factor that promoted the implementation of the intervention (50% of respondents). Good training was the most cited facilitator behind various types of personnel support (teacher, principal, and other administrator support; 58%, 54%, and 58% respectively). Further, quality training was indicated as an important facilitator to sustainability (endorsed by 21% of respondents) (Forman et al., 2009). In terms of sustainability, quality training

was rated behind only the visible impact of the program (37%) and personnel support (teacher and principal; 29% and 21% respectively).

In addition to the Forman et al.'s (2009) qualitative review, the role of training in the transportation of an evidence-based practice to a school context was examined using an experimental design by Shernoff & Kratochwill (2007). They studied the Incredible Years Classroom Management Program, which has demonstrated its success in efficacy-based contexts but there is limited support for the program in typical practice. Therefore, the purpose of the study was to examine whether training resources had an impact on the transportation of the program. Teachers in a videotape modeling were compared to teachers who received videotape modeling and additional consultation. In the videotape modeling condition, teachers were provided with a videotape and teacher manual while teachers in the consultation condition were provided three phone consultation sessions (45 to 60 minutes), in addition to the videotape modeling and manual. The consultation sessions adhered to a behavioral consultation framework and involved a problem identification interview, problem analysis interview, and a treatment evaluation interview. This sequence of consultation interviews was designed to identify classroom management strategies that were challenging to implement (problem identification), identify variables that facilitated adherence to the strategies (problem analysis), and determine the extent to which plans were successful in increasing adherence and discuss plans to enhance maintenance/generalization (treatment evaluation). Shernoff & Kratochwill (2007) examined whether the differences in training had an impact on teacher adherence, student behavior, and acceptability of the intervention. Results indicated that teachers in the consultation group more frequently used proactive

instructional strategies that were taught in the program and rated the intervention as more acceptable. In terms of student outcomes, students in the consultation group showed stronger positive results for ratings of externalizing problems, and demonstrated higher ratings of social competence. The findings of this study empirically supports the use of ongoing training (consultation) in promoting the transportability of an evidence-based intervention into a school context. Despite the encouraging effects of training on various outcomes (e.g. student outcomes, transportability, sustainability), research has also supported the use of ongoing performance feedback in order to promote treatment integrity.

Training to promote integrity. Interventionist training is significantly related to integrity of implementation (Noell et al., 1997; Witt et al., 1997; Noell et al., 2005), which in turn relates to the quality of intervention and increases the likelihood that the intervention will be successful in promoting desired outcomes. Noell et al., (1997) and Witt et al. (1997) began a line of research in schools that examined the role of training and specifically performance feedback on intervention plan implementation. Noell (2010) distinguishes intervention plan implementation by specifying that it focuses on the examination of implementation of an intervention or treatment plan in a natural context, while treatment integrity refers to experimental contexts. Performance feedback is defined as capturing, summarizing, and presenting data to teachers regarding their plan implementation, often in graphical form (Noell & Gansle, 2014).

Witt et al. (1997) utilized a single-case multiple baseline design to examine teachers' implementation of reinforcement-based, academic interventions for referred students. Teachers (n=4) were provided all of the materials necessary to implement the

intervention and were trained, in vivo, to implement the intervention. The study began with a day of complete intervention implementation prior to baseline such that all teachers exhibited 100% treatment integrity (based on the number of completed treatment steps). All teachers continued with a decreasing trend of integrity following baseline with a marked improvement following performance feedback, from a mean of 53% to 71% (percent of completed steps). The performance feedback consisted of a daily meeting during which the students' academic performance and teachers' intervention implementation was reviewed. The consultant specified missed treatment steps and suggested methods to enhance implementation. In addition to the increase in integrity as a result of the performance feedback, 3 of the 4 students demonstrated additional increases in academic performance following the increase in integrity.

The Noell et al. (1997) study was similar in its experimental design with the exception that they provided relatively limited didactic training prior to implementation (as opposed to the in vivo training provided in the Witt et al., 1997 study). The results were consistent across the two studies, which showed that intervention plan implementation was generally moderate to high in baseline and progressively deteriorated after a few days. This trend suggested that implementation of interventions may deteriorate rather quickly even with relatively extensive training and the provision of all the necessary materials (as in the Witt et al., 1997 study). During the performance feedback phase, teachers again demonstrated markedly higher levels of implementation.

These findings were replicated and extended in a randomized clinical field trial (Noell et al., 2005). In this study, teachers (n=45) engaging in behavioral consultation were randomly assigned to one of three follow-up conditions: 1) weekly follow-ups,

which consisted of a brief meeting between the consultant and the teacher that was structured as an abbreviated plan evaluation interview (i.e., discussed extent to which the plan was implemented, extent of student improvement, and any potential questions), 2) commitment emphasis, which included all of the elements of the weekly follow-up and included a social influence procedure at the final meeting (reviewed five points designed to enhance commitment to implementation), and 3) performance feedback, which was modeled after previous research and consisted of a review of intervention steps, graphing of student behavior, and graphing of intervention implementation. The three groups were compared on the extent to which teachers implemented the students' intervention plans as they were designed (based on the number of steps completed). The results of the study demonstrated a large effect size (1.3) relative to the baseline condition and demonstrated that the performance feedback was the strongest approach for improving and maintaining treatment integrity. The grand mean over three weeks was 77.1% integrity in the performance feedback condition, compared to 52.3% for the commitment condition, and 35% for the weekly follow-up condition. This finding highlights the importance of ongoing performance feedback as a specific component of training to promote quality implementation.

Summary. High-quality training is both conceptually and empirically related to important factors in successful intervention implementation. While training leads to improved implementation relative to no training, ongoing professional development (with performance feedback), contributes to even greater quality of implementation (Gansle & Noell, 2007; Ransford et al., 2009). High-quality training promotes the transportability of evidence-based interventions and also facilitates the sustainability of interventions

(Forman et al., 2009; Shernoff & Kratochwill, 2007). Further, training is known to contribute to greater treatment integrity, which as described above, is a critical feature of intervention implementation that relates directly to student outcomes (Durlak & DuPre, 2008; Noell & Gansle, 2006). Collectively, the provision of evidence-based, progress monitoring, treatment integrity, and interventionist training represent the empirical basis for high-quality reading intervention implementation in the context of RtI.

Summary of High-Quality Reading Intervention Implementation in RtI

Successful implementation of reading interventions in RtI requires that the underlying evidence is strong. Evidence-based reading interventions are distinguished by research findings that demonstrate they are likely to produce predictable, beneficial, and effective results (Al Otaiba et al., 2011; Denton et al., 2013; Goss & Brown-Chidsey, 2012; Wanzek & Vaughn, 2008). There are many different methodologies to define what consists of 'evidence-based' and therefore utilizing multiple databases to review intervention provide a comprehensive method for defining an evidence-based intervention (Slavin, 2008; Stoiber & Desmet, 2010). In addition to utilizing evidence-based interventions, progress monitoring is a necessity in the context of response-to-intervention (Brown-Chidsey & Steege, 2010). Specifically, curriculum-based measures have been shown to have strong psychometric properties, predictive utility, and can be used to inform instructional decision-making (Christ et al., 2013; Stecker, Fuchs, & Fuchs, 2005; Wayman et al., 2007). Third, the assurance of treatment integrity has been shown to contribute to positive student outcomes and enables the interpretation of evidence of effectiveness of the program (Durlak & Dupre, 2008; Noell & Gansle, 2006; O'Donnel, 2008). However, due to challenges in measuring integrity, it is recommended

that both indirect and direct measures be utilized (Hagermoser Sanetti & Kratochwill, 2014). Lastly, interventionist training is regarded as a critical component of high-quality intervention implementation. While training in general has been shown to promote positive outcomes, the use of ongoing performance feedback has been shown to promote transportability/sustainability and treatment integrity (Durlak & DuPre, 2008; Forman et al., 2009; Noell et al., 2005; Noell & Gansle, 2006).

As described above, these four components comprise the variable of interest, the 'what,' of high-quality reading intervention implementation in the context of RtI. From an implementation science perspective, factors that contribute to successful implementation, the 'how,' are of interest for empirical examination in order to determine factors that may contribute to successful (or unsuccessful) implementation (Ogden & Fixsen, 2014).

Challenges to Implementation of High-Quality Evidence-Based Reading Interventions

Despite the increasing prominence of evidence-based interventions, particularly in reading, it has been shown that, although our knowledge concerning interventions has increased, the frequency with which evidence-based interventions are implemented in school settings remains worryingly low (Forman & Burke, 2008). There is a gap between the existing knowledge of evidence-based interventions and their application in school practice settings (Forman & Burke, 2008). Many critics to the implementation of tiered systems of evidence-based interventions cite difficulty in translating the model from theory into large-scale practice (Garcia & Ortiz, 2008). The use of multiple evidence-based interventions requires a structural overhaul of schooling practices and changes in

traditional resource allocation. For example, schools must select, likely purchase, and implement high-quality interventions, provide training and build capacity to implement said interventions, and develop and implement progress monitoring procedures, all while ensuring that each component is implemented with integrity (Brown-Chidsey & Steege, 2010).

In reference to RtI, Fuchs, Fuchs & Compton (2012) write, the "one constant among the many variants of RtI is that, as an early intervention and prevention system, it is costly in time and resources. It requires assessments and interventions that educators rarely conducted a decade ago… because of its newness there are serious inefficiencies in its application" (p. 264). Therefore, many of the challenging demands of high-quality reading interventions in the context of RtI can be related to schools' financial capacity and achievement levels.

School Finance and Evidence-Based Reading Interventions

In the implementation science literature, a critical component that facilitates successful implementation is available financial resources at the organizational level (Fixsen et al., 2005). Not surprisingly, adequate funding may play a critical role in the implementation of high-quality reading interventions. Given the potential costliness of implementing evidence-based interventions, progress monitoring, monitoring integrity and training interventionists, there may be a link between financial resources and the quality of reading intervention implementation. While the actual costs of implementing evidence-based reading interventions is unknown and may vary from school to school, a description of the actual costs associated with implementing sample programs/tools (e.g.,

reading recovery and AIMSWEB) is presented below, in order to illustrate the potential costs of quality reading intervention implementation in the context of RtI.

Cost of evidence-based interventions. It is apparent that successful implementation of RtI models require evidence-based interventions to be selected and implemented for reading. Considering the many components of reading, different grade levels, as well as other diverse characteristics, it is unlikely that one intervention will be sufficient to meet these diverse needs of all students. Accordingly, school personnel will need access to multiple interventions to address the range of students' skills and needs, which may become problematic considering the cost of the interventions.

For example, consider the cost of *Reading Recovery*, a program that has been shown to be effective based on reviews by the What Works Clearing House, Promising Practices Network, and the Johns Hopkins Best Evidence Encyclopedia. *Reading Recovery* requires an extensive collection of short books with each teacher needing a set of starter books that will grow over the years. The initial starter set of materials is available for an estimated $3,000 dollars for one teacher. Further, the intervention only applies to kindergarten and 1st grade, and other, equally expensive programs may be necessary to provide evidence-based interventions at higher grade-levels (Requirements and costs for the training of reading recovery, 2013). Despite these expenses, additional costs are incurred in the form of training.

Cost of training interventionists. As described above, it has been shown that the mere provision of materials is insufficient to promote implementation. Successful intervention implementation requires both the necessary materials and personnel to implement the interventions. Therefore, some minimal form of training is necessary to

teach the interventionist how to implement the intervention, and it is recommended that training include ongoing performance feedback. In the case of *Reading Recovery*, for example, the bulk of the costs come in the form of training. There are only 23 *Reading Recovery* training centers in the US (predominantly on the east coast), which is the only place that a teacher leader can be trained (*Reading Recovery* Ccouncil of North America). The teacher must have teaching experience (three years total and at least two at the primary level), demonstrate evidence of adaptability, problem solving ability, willingness to learn/apply skills, and should be selected by a screening committee in consultation with a teacher leader (Requirements and costs for the training of *Reading Recovery* teachers in Ohio, 2011). Training as a *Reading Recovery* teacher requires participation in 2 graduate level courses taught by a certified reading recovery teacher leader over a full academic year. The tuition associated is approximately $4,700, and the support to the regional training site for operating costs is $3,000 per teacher in training (Requirements and costs for the training of *Reading Recovery*, 2013). In addition to these costs, with relatively few training centers available, considerable transportation costs may add further expenses.

In total, this amounts to an estimated cost of $10,700 for training and materials (without transportation costs) for one teacher for one year; and this does not include costs of travel or the considerable ongoing professional development required by the program in future years. While they may vary somewhat by the specific training site (e.g. tuition fees), the training is rather costly, particularly when considering, again, that this intervention applies only to early literacy (kindergarten and 1st grade) (Requirements and costs for the training of *Reading Recovery*, 2013).

Reading Recovery represents one of the more expensive options for an empirically-based intervention, but it is very comprehensive in terms of its evidential support and the training sequence that is provided. There are, of course, other options in terms of packaged interventions, but they often have less extensive training and support. Another option would be to develop interventions based on empirically-supported principles. For example, general strategies such as direct/explicit instruction, peer tutoring, and repeated reading have been shown to be effective (Hattie, 2009). Despite the lesser 'up-front' costs of these more informal approaches, it will require additional costs in terms of staff resources (for the development of the intervention and the ensuing training). Accordingly *Reading Recovery* and self-developed interventions represent a range of options in terms of financial costs, level of training, and staff development that a school must consider.

Cost of progress monitoring. Similar to evidence-based interventions, progress monitoring requires schools to select, purchase, and implement appropriate tools and procedures. Schools must purchase materials, which further require considerable training in administration, data collection/entry, data analysis, and interpretation in the context of multiple intervention implementation (i.e. establishing rules that consider rate of improvement and levels that may merit moving between tiers and/or changes in instructional practices) (Shapiro, Zigmond, Wallace, & Marston, 2011).

As an example, consider AIMSWEB, a commonly used program that provides multiple reading assessments, and has been validated by the National Center for Progress Monitoring. The cost of subscription for one student in reading is $4.00 (includes assessment content, scoring tools, data management and reporting) (AIMSWEB reading).

Assume a school of 500 students and adhering to the traditional 80-15-5 model of service provision, this means 20%, or 100, of the students (assuming progress monitoring at tiers 2 and 3) will need a subscription. This totals $4,000 dollars for the subscription alone. In addition to the cost of materials, training services are also a significant expenditure. For example, the AIMSWEB basic training costs $3,500 per session (up to 30 participants) and the training to utilize AIMSWEB data to guide instruction is an additional $2,500 (again up to 30 participants). However, given the recommendation that training occur in an ongoing fashion, it may be necessary to receive coaching throughout the year, which is offered online (1 hour remote consultation) for $250 per person. Therefore, providing two in-service training sessions (for 30 participants) and two additional coaching sessions (for 15 participants) would cost $13,500 (AIMSWEB reading). In total, the training of 30 people to provide progress monitoring (with ongoing coaching for half of those trained) for 100 students, costs roughly $17,500 for one year.

Similar to *Reading Recovery*, AIMSWEB represents one of the more costly options for progress monitoring tools. However this cost comes with training, computer programs designed for progress monitoring, and an assortment of validated tools. Other options may be lesser in cost (e.g. intervention central or self-developed probes/norms), but they may only provide the probes themselves and do not include the training and computer programs to be used for data management and graphing. These additional responsibilities would have to be established and developed by school personnel and thus require costs in terms of staff resources. In this manner, AIMSWEB or cheaper alternatives represent a range of options in terms of financial costs, level of training, and staff development that a school must consider.

Cost of integrity. Treatment integrity is the final core component of intervention implementation in the context of RtI. There is not a clear-cut fiscal cost for the utilization of integrity measures. However, assuring integrity will have financial implications in that schools will need to delegate the responsibility of developing integrity measures, conducting integrity checks, and analyzing integrity data (Shapiro et al., 2011). This requires that the treatment components be operationalized so that they can be directly observed by independent observers and requires the time, training, and staff flexibility for personnel to conduct walkthroughs to verify the appropriate implementation of the intervention and its components.

Cost of staff. While the above descriptions represent the raw costs of implementation of each component of high-quality interventions, it is important to also consider the financial costs in terms of staff time. There is not a clear method to estimate these costs (or empirical research on the costs), but it is important to realize that even after providing materials and training, considerable 'man-power' is necessary to administer interventions, monitor progress, and collect integrity data. This additional time demand equates to financial cost in that someone will have to be paid to fulfill these roles, whether it comes in the form restructuring the roles of existing personnel or in terms of hiring specialists to serve these functions (e.g. reading specialists, literacy coaches, resource teachers, etc.).

Summary. Given the premium resources required to implement high-quality evidence-based reading interventions, it is less surprising that the frequency with which evidence-based interventions are implemented in school settings remains low (Forman & Burke, 2008). Most notably, the high financial costs may pose a considerable challenge

to many schools seeking to utilize empirically-based reading interventions (Hagermoser Sanetti & Kratochwill, 2009). In light of the significant role of financial cost in implementation, it is important to consider how school districts obtain funding, and it is of pressing empirical interest to determine how funding may contribute to the implementation of quality reading interventions in the school context.

School Funding

Since education is primarily a state responsibility, more than 90% of school funding comes from state and local sources (Epstein, 2011), which has been a relatively long standing issue in the United States, with calls for reform dating back to 1972 (Equity and Excellence Commission, 2013). At that time (1972), the Commission on School Finance issued a report titled *Schools, People, Money: The Need for Educational Reform*, which explored the effects of reliance on property taxes to fund schools. They found that many of the problems with education funding were the direct result of antiquated state funding formulas that were over-reliant on local property taxes. Since districts have traditionally drawn much of their revenue from local property taxes, those in high-wealth parts of a state were often funded more generously than districts in low-wealth areas (Epstein, 2011); accordingly, people living in property-rich districts could fund their public schools more liberally even at lower tax rates relative to residents in lower-income areas. For example if Town A has $100,000 in taxable property compared to Town B's $300,000, a 2 percent tax rate in Town B raises $6,000 per student but a 4 percent tax rate in town A will only yield $4,000 dollars per student (Epstein, 2011). This example is an oversimplification, but highlights the problematic nature of relying heavily on local

revenue sources for school funding because schools with less available funding may have limited capacity to provide high-quality reading interventions.

Due to this clear inequity, many states have reduced their reliance on local funding sources and increased their reliance on state sources in an attempt to better level the playing field. The historical inequities as a result of local funding reliance have led to changes in systems of funding, but funding remains predominantly controlled by the state. Accordingly, there is considerable variability in funding practices across states, and despite some reform, the link between locally-based school finance and per-pupil spending inequities is still strong, and remains a concern for many states and localities (Equity and Excellence Commission, 2013).

Given the reliance on local property tax revenue, the recession of 2007 created a fiscal crisis impacting school funding (Equity and Excellence Commission, 2013). Declining property taxes, massive job losses, and the deterioration of local/state labor markets triggered a substantial decline in many sources of education funding (Baker et al., 2014). In response to the crisis, the American Recovery and Reinvestment Act (ARRA) allocated $100 billion in stimulus funds to public schools in order to avoid major layoffs, cuts in essential educational programs/services, and maintain prior levels of financial support (before the fiscal crisis). Despite its intent, the stimulus funding merely served as a short-term stop gap, and states are now facing the long-term effects (must develop methods to offset the loss of federal funds). As the country emerges from the recession, the role of funding remains an issue in equity and fairness in the state and local systems (Baker et al., 2014).

Analyzing funding fairness. Per-pupil expenditure represents a basic index for the amount of spending used to support the students (Baker et al., 2014). It represents the cumulative total of federal, state, and local funding per-student and indicates the spending capacity that schools have to use on staff, programs, and in this case, high-quality reading interventions. Given the significant cost of implementing high-quality reading interventions, it is necessary to understand and examine the fairness of school funding procedures in order to determine if they systematically impact the ability to implement high-quality reading interventions.

In measuring funding fairness, it is important to utilize a measure of funding equity that includes within state variation, as this is a primary source of variation and potential inequity. Additionally, measures of funding fairness must take into account the variations in student demographics and variations in the costs of delivering equal educational opportunity. Specifically, a state system that ensures equal educational opportunity must account for additional needs generated by student poverty. Consistent with this principle, varying levels of funding are required to provide equal educational opportunities to children with different needs (Baker et al., 2014). Considering within state variability and student demographics are significant in determining the state funding systems that promote equitable funding and, therefore, would enable the provision of high-quality evidence-based reading interventions.

States can promote equity by allocating more funding to districts with higher concentrated poverty levels, which is known as a "progressive" system (Baker et al., 2014). A funding system that allocates less to districts with high concentrations of poverty is considered a "regressive" system. If the state allocates roughly the same

amount across districts with varying needs, it is known as a "flat" system (Baker et al., 2014). Concentrated poverty is significant because student and school poverty relates to a multitude of factors that increase the costs of providing equal educational opportunity (e.g. gaps in achievement, racial composition, student mobility). Given these known correlates, funding should increase relative to the level of concentrated student poverty in order to promote fairness. If school funding remains inequitable, a reduction in the achievement gaps between the enfranchised and disenfranchised is infeasible (Baker et al., 2014); the inequity is especially salient given that a significant function of high-quality evidence-based reading interventions is to provide services and remediate difficulties for students at risk.

In a national report on fairness of school funding, Baker, Sciarra, and Farrie (2014) document four methods to measure funding fairness by state. The first is funding level, which is the overall level of state and local revenue provided to school districts, with adjustments made to reflect differences in regional wages, poverty, economies of scale, and population density. The second measure, funding distribution, is a measure of how funding is distributed across districts within a state, relative to student poverty. It shows whether a state provides more or less funding to schools based on their poverty concentration. Effort, the third measure of fairness, involves differences in state spending relative to state fiscal capacity (i.e. the ratio of state spending to state gross domestic product). Lastly, coverage measures the proportion of school-age children attending the state's public schools, as compared with those not attending the state's public schools (e.g. parochial schools, private schools, and home schooling). Coverage is significant because the share of the state's students in public schools, and the median

household income of those students, is an important indicator of the distribution of funding relative to student poverty (i.e., takes into account when more affluent households simply opt out of public schooling) (Baker et al., 2014).

Funding fairness in the state of study. When examining school funding in the state of study, it is important to consider the sources of funding. Approximately 43% comes from local property taxes (from the 2012-2013 school-year), while the remaining financial sources include the state (45%), federal government (8%), and other local sources (4%). However, being that this study occurred within one state, the primary interest in terms of funding fairness involve measures that capture within state variability in funding. Therefore, for the purposes of this study, overall funding level (fairness measure number one) and effort (the third fairness measures) are not significant, as they do not involve within state variability. Representing sources of within state variability, funding distribution and coverage are the two measures of fairness that are of interest within this research study.

Funding distribution. Funding distribution is based on the predicted per-pupil funding presented across different levels of poverty (10 percent intervals from 0 to 30 percent, i.e. 0-10%, 10-20%, 20-30%, and 30% or more) (Baker et al., 2014). A progressive state would have a high ratio between high- and low-poverty districts as the poor districts (i.e., high-poverty) would be receiving more funding than wealthy districts (i.e., low-poverty). For example, if districts with 30% poverty or more received $5,000 per student and districts with 0-10% poverty received $4,000 per student, the ratio would be greater than 1 (1.2), which represents a progressive system. This is in contrast to a regressively funded state which would have a low ratio between high- and low-poverty

districts. The measure of funding distribution takes on greater significance in the context of the declining revenues and school aid cuts because state policy will play a significant role in how these cuts will be distributed across school districts (Baker et al., 2014). In this way, states could potentially make cuts that target high-poverty districts and make the overall funding system more regressive.

The state of interest has maintained a stable and flat funding distribution since 2007 (Baker et al., 2014). Meaning, their approach has not changed significantly over the past five years (stable); for example, in 2007 their high/low-poverty funding ratio was approximately 96% and in 2011 it was approximately 99%. A perfectly flat funding distribution would equal 100% and essentially mean that all children (independent of their poverty level) receive the same level of funding. In this case (based on 2011 data), schools at 30% poverty receive approximately the same amount of money than schools at 0% poverty. To put this in perspective, the state of study was in the middle range relative to other states, from most progressive to most regressive. The most progressive state in 2011 (Minnesota) provided $13,547 for students at 30% poverty and $10,546 for students at 0% poverty. The most regressive state in 2011 (Nevada) provided $7,712 for students at 30% poverty and $11,145 for students at 0% poverty (Baker et al., 2014).

Coverage. The second indicator, coverage, involves the degree of disparity between those within and outside the public education system, while considering the number of students who attend public or non-public schools. A higher number of students within the public school system requires that a greater effort be made to fund those schools and a higher concentration of poor students in the public system requires more effort and greater attention to a fair distribution of funds (more progressive).

Coverage data therefore includes the percentage of school-aged children enrolled in public schools and the ratio of household income between public and nonpublic students, with states being ranked on a combined index of these two measures (Baker et al., 2014).

Relative to other states, the state of study is again in the middle of the pack, in terms of overall coverage in 2011. However, it is important to note that this rank stems moreso from the relatively high percentage of students attending private schools; this number is coupled with a 120% private-to-public income ratio, meaning the median household income for families of students attending private schools is 20% more than the median household income for families of students attending public school (Baker et al., 2014).

It is important to consider this finding in the context of the state's "flat" funding distribution. To clarify, the coverage indicator highlights that a relatively large portion of children from more affluent families attend private schools. Meaning, the pool of students attending public schools is already at a relative disadvantage prior to considering the funding distribution (since the money from the more affluent families is not being pooled into the public realm), and this creates less state impetus to provide a greater overall funding level (Baker et al., 2014). Further, it is with this already slight disadvantage, that the state provides a flat funding distribution, which actually serves to create greater inequity. Essentially, the state provides the same amount of money for students independent of their poverty level, which becomes problematic because the remainder of per-pupil expenditure dollars will come predominantly from local sources. This will serve to maintain the inequitable funding that has been historically problematic (over-reliance on local funding), because once again, residents in areas of greater wealth

will be able to fund their public schools more generously even at lower tax rates relative to residents in lower-income areas (exactly as in the case described above) (Baker et al., 2014).

Funding inequity in the context of reading interventions. The description of school funding highlights a glaring challenge for the provision of high-quality reading interventions in schools. Essentially, states continue to finance public education through methods that have no demonstrable link to the cost of delivering rigorous academic standards to produce high achievement in all students, particularly English-language learners, students with disabilities, students who live in high poverty schools/districts, and students who live in remote schools and districts (Equity and Excellence Commission, 2013). The funding issue is particularly pertinent in the context of the large and growing population of poor students who are concentrated in high-poverty school districts. While there are cases of schools that spend large sums of money and attain poor results, this is the exception rather than the rule, and most commonly, schools in high-poverty areas are less funded and have lower achievement levels. Unfortunately, this finding is well supported in the context of reading achievement. United States children in low-poverty schools rank among the top achievers of the world while those in our highest-poverty schools are performing on par with children in the world's lowest achieving countries (Equity and Excellence Commission, 2013). The discrepant achievement across poverty levels is, in part, an indicator of our historical inability to fund schools in a manner that promotes equitable student outcomes.

What follows is a paradox in the context of the evidence-based intervention movement and the state distribution of funding. On the one hand, evidence-based

interventions are now part of the larger legislative aims of No Child Left Behind and the Individuals with Disabilities Education Improvement Act (2004). In the NCLB context, evidence-based reading interventions are intended to enable educational reform (Fuchs, Fuchs, and Stecker, 2010), which is meant primarily to close the achievement gap between traditionally enfranchised and disenfranchised groups. In the IDEA context, evidence-based reading interventions have now become part of an alternative special education eligibility system (IDEA, 2004). The use of evidence-based interventions are intended to lead to more meaningful disability identification by accelerating progress of many low achievers and thereby eliminating them from consideration for special education, which is especially relevant for students of disadvantaged backgrounds (Fuchs, Fuchs, and Stecker, 2010). Indicative of the now supreme importance of implementing high-quality reading interventions, consider that a failure to appropriately implement high-quality reading interventions can lead to incorrect diagnoses of student disabilities.

The legislative movements clearly place a newfound premium on the implementation of high-quality evidence-based reading interventions, yet schools are implementing them at surprisingly and alarmingly low rates (Forman & Burke, 2008). However, this finding becomes nothing short of commonsensical when considering the high cost of interventions and the structure of school funding. Specifically, schools in disadvantaged areas receive less funding than those in wealthier areas due to the dependence on local funding sources. Accordingly, the students of schools in poorer areas are unable to support the financial burden of the interventions that are designed to promote more equitable reading outcomes (who are ironically the primary target of the

legislative movements). Put succinctly, the promotion of high-quality evidence-based reading interventions is designed, at least in part, with students who are disadvantaged in mind; yet, these students are in schools that are not funded sufficiently to engender their implementation, and as a result the interventions are not reaching their intended target. This leaves our highly touted intervention movement doomed to perpetuate the larger schooling inequities that has faced our nation for decades.

Inequitable funding coupled with the impetus to provide remedial evidence-based interventions creates a paradox among schools with low reading achievement. Schools with the greatest need for the use of empirical interventions are provided fewer financial resources to do so. In light of this, it is of empirical interest to examine how school finances and school reading achievement levels contribute to the quality of intervention implementation in the context of response-to-intervention. Low achieving schools may require greater levels of funding to meet the needs of their students. Accordingly, it is necessary to examine whether school funding and reading achievement levels impact the ability to implement quality reading interventions.

Achievement Levels

In addition to fiscal resources, reading achievement levels play an integral role in the implementation of high-quality evidence-based reading interventions. In this study, achievement levels are characterized by the proportion of students within a school who are at or below a basic level on the state standardized test (i.e., below proficiency). Schools with large numbers of students below proficiency have greater needs in terms of remediation. Specifically, achievement levels can have indirect effects on intervention quality, such that low achieving schools have a concomitant need for more funding, or

the effects may be more direct, such that low achieving schools have a greater number of students in need of intervention, which may create challenges in adhering to recommended practices (Baker et al., 2014; Shapiro et al., 2011).

Indirect effects. Achievement levels can impact finances through adequate staffing, facilitation strategies (e.g. training, coaching), and the amount of resources required (Hagermoser Sanetti, & Kratochwill, 2009). For example, the RtI model can be conceptualized as a resource-allocation model such that 15% of students are at risk for reading failure and receive Tier 2 supports, while 5% demonstrate significant difficulties in reading and need more intensive Tier 3 supports. However, achievement levels do not necessarily fit this template. For example, consider a low achieving school, with 30% of students at risk for reading failure (i.e. needing Tier 2 supports) and 10% of students demonstrating significant difficulties in reading (i.e. needing more intensive Tier 3 supports). In terms of the provision of evidence-based interventions this would mean double the additional intervention materials and teachers requiring training, ($11,700 dollars for training/materials for one teacher for one year), and not to mention the additional staff that would be required to provide those interventions. When considering the costs of progress monitoring this would also potentially double the costs, with up to 200 students (in a hypothetical school of 500) requiring progress monitoring ($8,000 dollars in materials as opposed to $4,000), and additional costs in terms of staff training.

Direct effects. Not only can achievement levels indirectly impact intervention quality through variables related to finances, but achievement levels also play a potentially direct role in the appropriate implementation of interventions. While it has yet to be examined experimentally, low achievement levels may create challenges in meeting

recommendations regarding sufficient intervention duration and group size (Shapiro et al., 2011). Low achieving schools will have more students in need of tiered supports (i.e. intervention and progress monitoring); therefore, with so many students in need of additional intervention, schools may face logistical challenges in providing interventions at the recommended dosage and group size (4 days per week, approximately 30 minutes per session, with groups of 4-6 at Tier 2; 5 days per week, with greater than 30 minutes per session and groups of 1-3 at Tier 3; Brown-Chidsey & Steege, 2010).

Case example. Thirty-two percent of students are performing below basic on the fourth grade NAEP assessment, and with levels of proficiency being known to be much worse in areas of highly concentrated poverty, the challenges of low reading achievement levels may be more the norm than the exception. This notion has gone dramatically understudied, and the following case example provides an illustration of how demanding the provision of a multi-tiered intervention can be in the context of a low achieving school.

Larue is a school that is located in a district that received funding from the Office of Special Education Programs (OSEP) to identify and develop multi-tiered framework. Larue was specifically selected as it was considered a more challenging context to implement the model. Specifically, the school was comprised of 58% minority students, 92% were receiving free or reduced-price lunch and the percent of students reaching grade-level oral reading fluency benchmarks on DIBELS was as follows: 18% at first grade, 10% at second grade, 12 percent at third grade, and 2 percent in fourth grade. In this context, the model and provision of tiered services had to be modified considerably due to logistical challenges (Shapiro et al., 2011).

Due to the significant numbers of students performing below grade-level benchmarks, applying the original theoretical model (80-15-5 and using recommended groups sizes for intervention) would have required 18 adults to provide the small-group intervention to fourth graders alone (in a school of approximately 50 fourth graders) (Shapiro et al., 2011). Accordingly, district administrators and principals reassigned personnel and reconceptualized professional roles to maximize their limited resources. The overhaul required the creative and flexible use of Title I and special education staff, redefining of classroom teachers' instructional responsibilities (to include intervention), rearranging daily schedules (constrained by the availability of space, schedules, case-loads, contract-mandated teacher preparation time, and the need to allocate instructional time to all academic areas), and the creation of a new position (school literacy coach). This rearrangement allowed 90 minutes of daily core instruction and 120 minutes of small-group intervention time per week. Despite all this overhaul (to which staff members expressed concern and contention), the targeted group interventions (Tier 2) contained 12 students on average, and the more intensive intervention contained 6 students on average, both of which are well above the recommended group sizes (4-8 for group-based interventions and 1-3 for individualized interventions) (Shapiro et al., 2011). Further, the utilization of progress monitoring and data-based decision making was new and therefore, thousands of dollars of new instructional materials and technological tools were required to support the process. District personnel even began writing grants to obtain further funds to purchase data collection and data management systems and a preliminary set of standard protocol intervention materials.

The above description represents the complexity and logistical constraints of school reorganization necessary to implement a system of multi-tiered interventions at a low achieving school (based on low reading proficiency). These challenges represent unique obstacles for schools with large portions of low performing students (Baker et al., 2014; Shapiro et al., 2011). Even in the example above, with considerable restructuring made possible by extra resources (funding and support from university support), they were unable to adhere to the model recommendations. The challenges facing low achieving schools may be exacerbated in contexts without the additional funding and university support; accordingly, reading achievement levels is hypothesized to contribute to the schools capacity to implement interventions with appropriate duration and group size (i.e. intensity) (Baker et al., 2014; Shapiro et al., 2011).

Summary and Research Questions

The use of evidence-based interventions has recently become greatly emphasized within education and psychology, particularly in reading (Shernoff & Kratochwill, 2007). However, even with the strong support of the evidence-based practice movement, the question of how to translate the interventions into practice remains substantial, and is of considerable empirical interest (Ogden & Fixsen, 2014). Response-to-intervention represents a prominent model to facilitate the adoption of reading interventions in practice-based contexts, which can be characterized by 1) empirically-based interventions, 2) progress monitoring, 3) treatment integrity, and 4) a trained interventionist. Despite the underlying empirical support, the use of multiple evidence-based interventions requires a structural overhaul of schooling practices and changes in traditional resource allocation. Specifically, many of the challenging demands of high-

quality reading interventions in the context of RtI may be related to school financial capacity and achievement levels. Therefore, the purpose of this study is to examine how financial capacity and achievement levels contribute to the implementation of high-quality reading interventions in the context of RtI. Specifically:

1. Does per-pupil expenditure and student's reading proficiency predict the quality of evidence-based interventions, progress monitoring tools, treatment integrity measures, and trained at Tier 2?

    1a. Does per-pupil expenditure and student's reading proficiency predict the use of quality of evidence-based interventions at Tier 2?

    1b. Does per-pupil expenditure and student's reading proficiency predict the use of quality progress monitoring tools at Tier 2?

    1c. Does per-pupil expenditure and student's reading proficiency predict the use of quality treatment integrity measures at Tier 2?

    1d. Does per-pupil expenditure and student's reading proficiency predict the use of quality interventionist training at Tier 2?

2. Does per-pupil expenditure and student's reading proficiency predict the quality of evidence-based interventions, progress monitoring tools, treatment integrity measures, and trained at Tier 3?

    2a. Does per-pupil expenditure and student's reading proficiency predict the use of quality of evidence-based interventions at Tier 3?

    2b. Does per-pupil expenditure and student's reading proficiency predict the use of quality progress monitoring tools at Tier 3?

2c. Does per-pupil expenditure and student's reading proficiency predict the use of quality treatment integrity measures at Tier 3?

2d. Does per-pupil expenditure and student's reading proficiency predict the use of quality interventionist training at Tier 3?

3. Does student's reading proficiency predict the intensity (based on dosage and group size) of intervention implementation at Tier 2?
4. Does student's reading proficiency predict the intensity (based on dosage and group size) of intervention implementation at Tier 3?

CHAPTER THREE

Methods

The present study was part of a larger study designed to assess how schools (grades K-8) in a Midwestern state were utilizing response-to-intervention to manage the specific learning disability rule that allows the use of a process of systematic evidence-based interventions. As part of the larger study, a survey was administered two consecutive years (2011-2012 and 2012-2013) to track longitudinal developments, however, for the purposes of the present study only data from the 2nd year were utilized. The survey was comprised of 66 questions, of which a subset of questions was used for this study (described in detail below). The survey questions were used to capture the: 1) indicators of quality of reading intervention implementation, which includes evidence-based reading interventions, progress monitoring, treatment integrity, and trained interventionist, and 2) indicators of intervention intensity, which includes intervention dosage and group size. Accordingly, the survey data was used to examine how reading achievement levels and school finances impact the quality of intervention implementation and intervention intensity.

Instrumentation

Data were collected for this study using the abovementioned survey and the state database. Each of these tools is described below.

Survey. The original survey was comprised of 66 questions designed to assess how school personnel were utilizing response-to-intervention to manage the specific learning disability rule that allows the use of a process of systematic evidence-based interventions. Given the focus of the present study on evidence-based reading

interventions and how they were implemented, only a subset (18) of these 66 questions was used (see Appendix A for the survey questions and response options). A total of 14 survey questions were used to address the question of whether finances and achievement levels contribute to the quality of intervention implementation in the context of RtI. Specifically, survey questions 1-2 addressed the use of evidence-based interventions, questions 3-6 addressed the use of progress monitoring tools, questions 7-8 addressed the use of treatment integrity measures, and questions 9-14 addressed the use of a trained interventionist. To address the role of achievement levels and intervention intensity, four survey questions were used. Specifically, respondents were asked about the group sizes at tiers 2 and 3 (questions 15-16) and the duration of intervention based on the frequency of intervention sessions and duration of each session (questions 17-18).

State database. Data for the two independent variables (i.e., achievement levels and financial capacity) were gathered through the state database, which is a publically-accessible database in which state, district, and school level data are reported for four categories: 1) academic data (based on the state standardized assessment), 2) attendance/behavior (e.g. attendance, suspension, expulsion), 3) programs, staff and money (e.g. AP courses, teacher qualifications, revenue/cost per member), and 4) student demographics (e.g. ELL, racial/ethnic minority, students with disabilities, gender, migrant, economic disadvantage). The data used for the purposes of this study included school-level demographic data, academic data, and financial data. To provide a snapshot of the school characteristics, school-level percentages of student demographics were obtained. For the independent variables, achievement levels were based on reading proficiency levels on the state standardized test, while finances were based on the cost-

per-member (i.e., per-pupil expenditure) from the 2012-2013 academic year (the same year as the survey administration).

Data Collection

In this section, the coding for all study variables is explained. The independent variables include per-pupil expenditure and reading achievement levels (as defined by reading proficiency). As explained earlier, the per-pupil expenditure and reading proficiency data were gathered from the state database. The dependent variables included evidence-base interventions, progress monitoring, treatment integrity, trained interventionist, and intervention intensity (defined by group size and intervention duration). All of the data for the dependent variables were collected from the survey.

Per-pupil expenditure. Cost-per-member (referred to as per-pupil expenditure hereafter) was available publicly through the state website. Data were only available at the district level, and although school-level data would be preferred, many schools were the only elementary school in the district, and among districts with multiple schools represented in the survey, it would be expected that per-pupil expenditures were similar since schools represented similar grade levels. However, this is discussed as a limitation to the study.

Reading proficiency. Reading proficiency was based on the state administered standardized accountability test. The state accountability test was designed to measure the state standards and provide information about student attainment of subject-area proficiency (reading, math, science, language arts, writing, and social studies). Students in grades 3-8 and 10 completed the reading test. Students' scores were categorized into one of four performance levels: minimal, basic, proficient, and advanced. Minimal and

basic signifies that the students were scoring below the state standard benchmarks for grade-level reading. Data were collected at the school-level based on the percentage of students who scored at the minimal and basic levels. In order to capture achievement levels, if the percent of students at minimal and basic were to be highly correlated (>.80), they would be combined and treated as a single variable (percent of students below proficiency). If they were not strongly correlated (<.80) they would be treated as two separate predictor variables to better capture the role of achievement levels in predicting high-quality reading interventions and intervention intensity.

Evidence-based reading interventions. To address the use of evidence-based reading interventions, respondents were asked what interventions were used for students receiving Tier 2 and Tier 3 instruction for reading/language arts. Since the focus of this study is on reading, language arts response options were not coded in this analysis. At both Tier 2 and Tier 3, respondents were provided with response options representing a number of common interventions and they were able to select any that may apply. What Works Clearinghouse, Promising Practices Network, and Johns Hopkins Best Evidence Encyclopedia were used to code the interventions to determine the quality of the evidence-base. These three organizations were selected because they provide three different robust methods of review (mix of criteria for research controlled trials, quasi-experimental designs, and meta-analyses) and also provide descriptive summaries of evidence (e.g. no evidence, limited evidence, promising evidence, or strong evidence). Collectively they represent the best available research evidence (i.e. evidence-based interventions), and therefore their use should promote the implementation of effective

practices, as these interventions have been shown to utilize reliable and valid methods that lead to successful intervention outcomes in reading.

Specifically, Promising Practices network offers the following ratings: proven, promising, or other (have not been reviewed yet). Johns Hopkins Best Evidence Encyclopedia provides ratings of: strongest evidence, moderately strong evidence, and limited evidence. However, What Works Clearinghouse (WWC) provides a rating on multiple indicator domains (e.g. fluency and comprehension) with the following potential ratings (from strongest to weakest effects): ++ (very positive), + (positive), +/- (mixed), 0 (negligible), - (negative), -- (very negative). In order to provide one indicator, the modal rating of the indicator domains was used as the overall indicator for a particular program, and for instances of multiple modes, the median was used.

Based on the ratings from these three databases, an overall rating for each school was created based on the average rating of the interventions selected by that school and the number of interventions being used at each school. First, the average rating of each intervention was examined across databases (on a -1 to 1 scale). By database, being rated positively (i.e. receiving a 1) consisted of a modal rating of ++ or + in What Works Clearinghouse, a rating of moderately strong or strongest in Johns Hopkins Best Evidence Encyclopedia, and a rating of promising or proven in the Promising Practices network. This method was selected because it does not penalize an intervention for being rated in only one database while also accounting for cases in which the intervention was rated in multiple databases. There were also cases where general evidence-based practices (e.g. drill/flashcards, partner reading, and peer tutoring) were used; because this lacks a degree of specificity, these practices were coded as a .5. If a program was not

rated in any of the three databases, a search for the intervention title was conducted in PsychInfo and Academic Search to examine potential empirical research underlying the intervention. If no research was found they were considered not to be evidence-based and coded as a 0. If empirical research articles were found, the IES practice guide (criteria used by What Works Clearinghouse, 2010) was used to code the study (preference given to meta-analyses or research syntheses). General criteria ranged from minimal, moderate, to strong (0, .5, and 1 respectively) on the following components: validity, effects on relevant outcomes, relevance to scope, and relationship between research and recommendations (the remaining categories are not applicable to the individual coding, e.g. panel confidence, expert opinion). Further, if the manufacturer of the intervention were to provide the only empirical evidence available, the maximum effectiveness rating was moderate evidence (.5) due to the lack of independent replication.

After coding each intervention, to obtain an overall school-level indicator of the use of evidence-based interventions, the average rating of the interventions being used (for each school) was multiplied by the number of interventions being used. This coding scheme created continuous rating of evidence-based interventions based on the empirical support underlying the interventions and the availability of multiple interventions to address specific student needs.

Progress monitoring. The coding of the progress monitoring data was based on two survey questions asking which progress monitoring tools were being used. One question was for tools used at the Tier 2 level and the other question was the same but for the Tier 3 level. Progress monitoring tools were coded on a 0-1 point scale based on

review by the National Center for Progress Monitoring. The 0-1 point coding scale is consistent with the Center's approach, which specifies the following evidence-levels: data unavailable or unconvincing evidence (0), partially convincing evidence (.5), and convincing evidence (1). If a measure had been reviewed by the National Center for Progress Monitoring, it indicated the measures psychometric properties have been empirically examined. However, sub-skill mastery measures could be used for progress monitoring purposes but are not empirically validated by an external source, and therefore, were coded as partially convincing. Response options coded as a 0 were those that were intended to be used as a screening measure or that were part of an intervention package and did not necessarily represent a validated progress monitoring tool (e.g. Imagine Learning, Compass Learning, and Read 180).

In addition to the use of a psychometrically sound progress monitoring tool, the frequency of use was an important consideration of quality because frequency impacts the reliability and validity of the slopes of student progress, which are used for decision-making purposes. The consideration of progress monitoring frequency was based on two survey questions, one for Tier 2 interventions and another for Tier 3 interventions. The questions asked how often progress monitoring was conducted for reading, with the following response options: daily, 4 times per week, 3 times per week, 2 times per week, weekly, every two weeks, and once per month. To create an overall composite, the frequency of progress monitoring was multiplied by the rating of the quality of the progress monitoring tool, which created a continuous indicator of quality progress monitoring. This created an overall coding scheme from 0 to 5, with frequency coded in

the following manner: 5 (daily), 4 (4 times per week), 3 (3 times per week), 2 (2times per week), 1 (weekly), .5 (every two weeks), and .25 (once per month).

Treatment integrity. To address the use of intervention integrity measures, respondents were asked two questions, one at the Tier 2 level and one at the Tier 3 level. Respondents were provided the following options: checklist, video-taping, peer feedback, outside monitoring, and other (write-in). The highest quality of integrity monitoring was to use a combination of direct and indirect measures. However, due to the known importance of external feedback, direct methods with feedback were given preference. The following coding scheme was developed to capture these gradations in quality. Five points were given to schools using a direct method with feedback (e.g. peer feedback or outside monitoring) and an indirect method (e.g. checklist). Four points were given to schools using a direct method without feedback (e.g. videotaping) and an indirect method. Three points were for schools using only a direct method with feedback and two points were for schools using only a direct method without feedback. One point was for schools using only an indirect method, and zero points were provided if schools were not using any method to monitor integrity. Therefore, quality treatment integrity was coded on an ordinal scale from 0 to 5.

Trained interventionist. A total of six questions were used to address the use of a trained interventionist, three at the Tier 2 level and three at the Tier 3 level. The first two questions (at each tier) were used to assess who was implementing the intervention and who was trained to implement the interventions. The third question (at each tier) asked more specific information regarding the type of training provided, asking how the school personnel developed the capacity to implement interventions used at Tier 2 and

Tier 3. For this question, response options were as follows: ongoing professional development, in-service day(s), materials introduced at a faculty meeting, materials introduced at a team meeting, materials provided with expectation to review and implement, and no training provided to date. In coding the use of a trained interventionist, the gold standard was to provide in-service training in addition to on-going professional development, due to the importance of robust initial training with continued coaching/feedback. Therefore, the following coding scheme was developed to capture the distinctions in quality training. Four points were provided to schools who utilized ongoing professional development and in-service day(s), three points were provided for only on-going professional development, two points were provided for only in-service day(s), one point was provided for any other method (materials introduced at a faculty meeting, team meeting, and with the expectation to review and implement), while zero points was provided for no training. Therefore, the final coding was on an ordinal scale from 0 to 4 based on the overall quality of training.

Intensity of intervention. Intensity of intervention was based on a combination of group size and intervention dosage. Based on the coding scheme described below, a continuous indicator of intensity was created to capture the number of minutes of intervention provided per week per student at each tier.

Intervention dosage. The intervention dosage indicator was used to capture the weekly duration of intervention implementation. It was based on a two part question for both the Tier 2 and Tier 3 levels. The question asks how often reading interventions were typically provided to students with response options in two parts. The first part indicated the frequency of the sessions with the following options: 1, 2, 3, 4, 5, or 5+; the second

part indicated the duration of each session in minutes with the following options: less than 15, 15-30, 31-45, and more than 45. The frequency and duration per intervention session were used to create a weekly total of intervention duration by multiplying the number of sessions per week by the length of each session. In coding the frequency of each session, 5+ was coded as a 6, and in coding the duration of each session, the higher number of the response option range was used; specifically, less than 15 minutes was coded as 15, 15-30 was coded as 30, 31-45 was coded and 45, and more than 45 was coded as 60. Multiplying the frequency of intervention sessions (per week) by the duration of each session created a minutes-per-week measure of intervention dosage.

Group size. Two survey questions about group size were used to capture group sizes at Tier 2 and Tier 3 intervention levels. The questions asked how Tier 2 and Tier 3 reading interventions were typically provided to students with the following response options: individualized instruction, group of 2-3 students, group of 3-5 students, and group of 6 or more students. Group size was used in conjunction with dosage to create an indicator of intensity per student and, therefore, the following coding scheme was used to code the group size response options: 1 (individualized), 2.5 (group of 2-3 students), 4 (3-5 students), and 6 (6 or more students).

Composite. Based on the coding of intervention dosage and group size, the combination of these two variables was used to create an overall indicator of intervention intensity. Specifically, the dosage per week was divided by the group size to create a continuous measure of intensity based on the number of minutes per week per student.

Procedures

Survey distribution. School districts (of the 156 schools) were sent a request for participation directly from the state Department of Public Instruction. Districts were instructed to contact each school principal in their district, via email, who was then specifically requested for participation. The request for participation document outlined the purpose, goals, and process for completion of the survey. Districts were given approximately two weeks to forward the letter of participation to schools within the district. Another two weeks were allotted to make contact with individual schools regarding participation. Participating schools were asked to complete a computerized survey (using Qualtrics) between September 24th, 2012 and December 3, 2012 and encouraged to collaborate with others to ensure accuracy of the survey responses. Participating schools were sent reminder emails during the survey window specifying the end date for the survey. For schools that did not respond to the emails or fill out the survey, direct phone calls to the principals were made to request participation.

Design and Data Analysis

Descriptive analysis. All of the data were entered in SPSS (version 22) for analysis. First, descriptive analysis of all independent and dependent variables was conducted. Descriptive data included means and standard deviation. Additionally, a correlation matrix for all independent and dependent variables was conducted.

Covariate analysis. In addition to providing descriptive statistics, analysis of the data for important group differences to identify potential covariates was conducted. Specifically, the correlation matrix was used to determine if there were significant relationships between the percent of students receiving free/reduced lunch (i.e.,

economically disadvantaged), racial/ethnic minority, English Language Learners, and students with disabilities and any of the dependent variables. Variables identified as significant were included as predictors in the analyses for the research questions delineated below.

Research questions. A combination of simple regression and ordinal logistic regression was used to answer research questions one and two. Before presenting the analyses, the research questions are delineated below.

Research question #1. The primary research question of interest was whether the use of high-quality reading interventions (i.e., the quality of evidence-based interventions, progress monitoring tools, treatment integrity measures, and use of trained interventionists) is predicted by per-pupil expenditure and students' reading proficiency levels at Tier 2. More specifically, the following research questions were addressed in this analysis:

1a. Does per-pupil expenditure and student's reading proficiency predict the use of quality of evidence-based interventions at Tier 2?

1b. Does per-pupil expenditure and student's reading proficiency predict the use of quality progress monitoring tools at Tier 2?

1c. Does per-pupil expenditure and student's reading proficiency predict the use of quality treatment integrity measures at Tier 2?

1d. Does per-pupil expenditure and student's reading proficiency predict the use of quality interventionist training at Tier 2?

Research question #2. The second primary research question of interest was whether the use of high-quality reading interventions (i.e., the quality of evidence-based

interventions, progress monitoring tools, treatment integrity measures, and use of trained interventionists) was predicted by per-pupil expenditure and students' reading proficiency levels at Tier 3. More specifically, the following research questions were addressed in this analysis:

2a. Does per-pupil expenditure and student's reading proficiency predict the use of quality of evidence-based interventions at Tier 3?

2b. Does per-pupil expenditure and student's reading proficiency predict the use of quality progress monitoring tools at Tier 3?

2c. Does per-pupil expenditure and student's reading proficiency predict the use of quality treatment integrity measures at Tier 3?

2d. Does per-pupil expenditure and student's reading proficiency predict the use of quality interventionist training at Tier 3?

A regression based approach was appropriate for these questions because per-pupil expenditure and achievement levels (based on percent of students at minimal and percent of students at basic) were treated as continuous independent variables. However, the dependent variables differed in terms of being continuous or ordinal. Therefore, for questions 1a, 1b, 2a, and 2b, simple regression was used because the dependent variables were continuous (quality of evidence-based interventions and progress monitoring). Assumptions to be tested for these questions include, linearity, normality, and homoscedasticity; additionally, multicollinearity was examined to ensure predictors were sufficiently orthogonal. For questions 1c, 1d, 2c, and 2d the dependent variables were ordinal (treatment integrity and trained interventionists), and therefore, ordinal logistic regression was used. For these questions, assumption checking was not necessary since

logistic regression does not require adherence to any assumptions about the distributions of predictor variables (Tabachnick & Fidell, 2007); however, due to the sensitivity of logistic regression to high correlations among predictor variables, multicollinearity was examined to ensure orthoganality.

Due to the differing forms of regression that were being utilized, each dependent variable was examined independently. This allowed an examination of per-pupil expenditure and achievement levels (percent of students at minimal and percent of students at basic) and their impact on high-quality reading interventions in the context of RtI. These analyses were conducted separately at the Tier 2 and Tier 3 level, and accordingly, the Tier 2 level was considered its own family of tests (with the use of a Bonferroni adjustment to maintain a family-wise alpha of .05), and the Tier 3 level was also considered its own family.

In addition to analyzing the effect of finances and achievement levels on indicators of intervention implementation quality, a secondary purpose of the study was to also understand the effect of achievement levels on intervention intensity. The independent variable was based on reading proficiency (conceptualized equivalently as in research questions 1-2). The specific research questions are presented below:

Research question #3. Does student's reading proficiency predict the intensity (based on dosage and group size) of intervention implementation at Tier 2?

Research question #4. Does student's reading proficiency predict the intensity (based on dosage and group size) of intervention implementation at Tier 3?

Simple regression was used for both research questions, which was appropriate because both questions had one continuous dependent variable. Assumptions to be tested

for these questions include, linearity, normality, and homoscedasticity; additionally, multicollinearity was examined to ensure predictors were sufficiently orthogonal. In this context, examining the role of achievement levels was used to determine if reading proficiency impacts the intensity of reading intervention implementation in the context of RtI.

CHAPTER FOUR

Results

Respondents

The survey was distributed to school personnel at 156 schools between September 24th, 2012 and December 3, 2012. With 92 respondents, the response rate was 59.0%. From the initial sample, eight more schools were removed leaving a final sample of 84 respondents. Five schools were removed due to unavailable data (four missing academic data and one missing per-pupil expenditure), two schools were removed since they did not represent the grade range requested (one high school and one preschool), one was removed because the respondent represented an online school, and lastly, one respondent did not complete the survey. The only demographic data collected on the respondents was their position. The majority of respondents were school principals (n=70; 83%), followed by other personnel (e.g., district administrator, co-administrator, RtI coordinator) (n=7; 8%), directors of special education (n=3; 4%), special education teachers (n=3; 4%), and classroom teachers (n=1; 1%). Respondents represented schools from 53 different school districts, which is 13% of the districts in the state of interest.

Description of Schools and Student Demographics

Schools. Seventy-nine (94%) respondents represented elementary schools, 3 (4%) represented middle schools, and 2 (2%) represented both elementary and middle schools (i.e., two separate schools). The mean size of the student population across schools was ~404 students; however they varied considerably in size. Specifically, the standard deviation of school size was ~174, with a range from 30 students to 802 students. As shown in Table 1, the majority of schools served students in grades K4-5 (n=27; 33%),

followed by PK-6 (n=11; 13%), K4-6 (n=7; 8%), K4-8 (n=7; 8%), K4-4 (n=6; 7%), while the remaining grade ranges had 5 or fewer schools (6% or less).

Table 1

Frequency of Sample Schools by Grade Range

Grade Range	Frequency	Percent
PK-4	2	2.4
PK-5	5	6.0
PK-6	11	13.3
K4-3	4	4.8
K4-4	6	7.2
K4-5	27	32.5
K4-6	7	8.4
K4-8	7	8.4
K5-4	2	2.4
K5-5	1	1.2
K5-6	1	1.2
K5-8	1	1.2
K5-12	1	1.2
1-4	1	1.2
1-5	1	1.2
1-8	1	1.2
2-4	1	1.2
3-5	1	1.2
3-6	1	1.2
5-8	2	2.4
6-8	1	1.2

Student Demographics. Demographic data regarding the student population of the schools were obtained through the state public database. Table 2 includes the racial/ethnic breakdown of the students across all 84 schools. To provide context, state level averages for elementary schools are also presented.

Table 2

School-level Descriptive Data for Student Racial/Ethnic Diversity

	Mean	Std. Deviation	Minimum	Maximum	State Average
Native Am.	0.72%	1.05%	0.00%	6.40%	1.28%
Asian	3.48%	5.21%	0.00%	29.41%	3.76%
African Am.	3.11%	10.39%	0.00%	95.68%	10.58%
Hispanic	5.80%	5.15%	0.00%	26.61%	11.69%
PI	0.06%	0.16%	0.00%	1.06%	.09%
2 or More	1.978%	1.83%	0.00%	7.14%	2.55%
White	84.87%	12.99%	2.88%	97.81%	69.95%

Note: The school level data are presented as percentages of the total student population.

As can be seen in Table 2, the student population was predominantly White (85%), followed by Hispanic (6%), Asian (3%), African American (3%), 2 or more races (2%), Native American (>1%) and Pacific Islander (PI) (>1%). Despite the majority of schools having been predominantly White in racial/ethnic background, there was variability in the racial/ethnic breakdown within schools, particularly in terms of the Asian, White, and Hispanic populations. Notably, there were schools with smaller White populations (as low as 2.88%) and larger Hispanic, Asian, and African American populations (maximum of 29.41%, 26.61%, and 95.68% respectively). Relative to state averages, the proportion of African American students and Hispanic students within the

sample was lower. This is likely because the sample was under-represented in terms of schools from urban districts.

In addition to racial/ethnic diversity, the percent of students who were of free or reduced lunch status (FRL), English Language Learners (ELL), male, migrant, and with disabilities is reported in Table 3.

Table 3

School-level Descriptive Data for Student Demographics

	Mean	Std. Deviation	Minimum	Maximum	State Average
FRL	32.75%	15.69%	0.00%	81.84%	45.13%
ELL	3.27%	4.13%	0.00%	20.43%	7.13%
Male	52.00%	2.66%	46.64%	60.58%	51.71%
Migrant	0.04%	0.15%	0.00%	0.85%	.07%
Disability	13.43%	4.24%	0.68%	26.83%	14.17%

Note: The school level data are presented as percentages of the total student population.

Schools in the sample varied considerably in the percent of students who were receiving free or reduced lunch (mean of 33% and standard deviation of 16%). Approximately 13% of students in the sample had disabilities, 3% were of ELL status, and there were more males (52%) than females (on average). Lastly, there was a small percentage of migrant students (.04%). Similar to the racial/ethnicity breakdown, relative to state averages, the proportion of students with disabilities, ELL, and FRL status within the sample was lower. This is likely because the sample was under-represented in terms of schools from urban districts.

Descriptive Analyses

All of the data were entered in SPSS (version 22) for analysis. Descriptive analyses (means and standard deviations) of all independent and dependent variables are presented in Table 4 and Table 5, respectively. Descriptive data regarding the dependent variables are presented in greater detail below with comparisons between Tier 2 and Tier 3 also described.

Independent variables. The two independent (predictor) variables in this study were reading achievement level and per-pupil expenditure. Achievement levels were defined by the percent of students who obtained a score of minimal or basic in reading on the statewide accountability assessment. The mean percent of students scoring minimal in reading on the state standardized test was 21% (standard deviation = 7%) and the mean percent of students scoring basic in reading on the state standardized test was 38% (standard deviation = 5%). The range of the percent of students scoring minimal in reading on the state standardized test was from 6% to 37%, while the range for the percent of students scoring Basic in reading on the state standardized test was from 26% to 55%.

Per-pupil expenditure (PPE) was the second predictor variable in this study. The mean per-pupil expenditure was equal to $12,395 (standard deviation = $1,538). The minimum per-pupil expenditure was $9,716 and the maximum per-pupil expenditure was $22,801. In comparing sample results to the state, the per-pupil expenditure and percent Basic were comparable, however more students were performing in the Minimal range in the state relative to the sample.

Table 4

Means and Standard Deviation for Independent Variables

	Mean	Std. Deviation	Minimum	Maximum	State Average
Minimal	21.42%	7.22%	6.40%	37.14%	29.65%
Basic	38.42%	5.18%	25.93%	54.79%	36.00%
PPE	$12,409.47	$1,550.77	$9,716	$22,801	$12,512

Note: Minimal = percent of students scoring minimal; Basic = percent of students scoring at basic; PPE = per-pupil expenditure

Dependent Variables. In total, there were five dependent variables examined in this study: 1) evidence-based interventions (EBI), 2) progress monitoring (PM), 3) treatment integrity (Integrity), 4) interventionist training (Training), and 5) intervention intensity (Intensity). Each of these variables is described further below (see Table 5 for the mean and standard deviation for each dependent variable).

Table 5

Means and Standard Deviation for Dependent Variables

	Mean	Std. Deviation	Minimum	Maximum
T2EBI	1.44	0.98	0	4
T2PM	0.76	1.02	0	5
T2Training	2.70	1.57	0	4
T2Integrity	1.07	1.78	0	5
T2Intensity	63.28	55.58	0	225
T3EBI	1.11	0.99	0	3.17
T3PM	0.88	1.22	0	5
T3Training	2.31	1.67	0	4
T3Integrity	1.57	1.85	0	5
T3Intensity	113.29	83.28	0	270

Note: T2EBI = Tier 2 evidence-based intervention rating; T2PM = Tier 2 progress monitoring rating; T2Training = Tier 2 interventionist training rating; T2Integrity = Tier 2 integrity measures rating; T2Intensity = Tier 2 intensity (minutes per session per student); T3EBI = Tier 3 evidence-based intervention rating; T3PM = Tier 3 progress monitoring rating; T3Training = Tier 3 interventionist training rating; T3Integrity = Tier 3 integrity measures rating; T3Intensity = Tier 3 intensity (minutes per session per student)

Evidence-Based Interventions. Based on the ratings from three databases (the What Works Clearinghouse, Promising Practices Network, and Johns Hopkins Best Evidence Encyclopedia), an overall rating for each school was created based on the average evidence rating across the three databases of all the interventions used by each school. As explained in the methods section, the criteria for an intervention to receive a value of 1 in each database was as follows: a modal rating of ++ or + in What Works Clearinghouse, a rating of moderately strong or strongest in Johns Hopkins Best Evidence Encyclopedia, and a rating of promising or proven in the Promising Practices

network. Interventions that were not rated within a database were denoted by NR (not rated). Interventions that were not rated in any of the three databases were searched in PsycInfo and Academic Search and then coded in accordance with IES guidelines (as described in the Method section).

As can be seen in Table 6, the following interventions were rated as 'evidence-based' (i.e., score of 1): Read 180, Reading Recovery, and Sound Partners, indicating strongly positive outcomes across all databases that provided ratings. The following interventions were rated as moderately evidence-based (.67 or .50): Accelerated Reader, Corrective Reading, Early Intervention in Reading, Reciprocal Teaching, Lexia, PALS, and general practices (Drill/Flashcards, Guided Reading, Individualized Instruction/Tutoring, Partner Reading, and Peer Tutoring). The relative frequency of moderately evidence-based interventions indicated a mix of ratings, such that the different databases provided mixed support for the effectiveness of the interventions (e.g. rated as positively effective in one database but not in another). The following interventions were reviewed by one or more of the three databases and were rated as not having an evidence-base (at this time): Orton-Gillingham, Read Naturally, Readers' Theatre, Reading Mastery, Wilson Reading System, and Voyager. Lastly, the following interventions were not rated in any of the three databases: Compass Learning Odyssey, Leveled Literacy Intervention, SOAR to Success, Imagine Learning, and Early Reading Empowerment. Upon search in PsychInfo and Academic Search no articles meeting IES guidelines were found (dissertations were excluded), and accordingly, they were rated as having no underlying evidence-base at this time. Descriptive data regarding the

implementation of evidence-based interventions as organized by tier is presented in the following section.

To obtain an overall school-level indicator of the use of evidence-based interventions, the average rating of all the interventions being used (for each school) was multiplied by the number of interventions being used at each school. The mean number of interventions used by each school was 3.94 interventions (standard deviation = 2.45), with schools using a range from zero to as many as nine interventions. The average rating of the interventions for each school was .36, ranging from zero to one. Overall, the mean rating of Tier 2 evidence-based reading interventions was 1.44 (standard deviation = 0.98). Interpreting this composite, schools typically had multiple interventions available (approximately four), but the underlying empirical support of those interventions was moderate at best. The most commonly used interventions included Guided Reading (n=59), Leveled Literacy Intervention (n=56), and Tutoring (n=36). See Appendix B for frequency data on all interventions at tiers 2 and 3.

At Tier 3, the mean number of interventions used by each school was 2.63 (standard deviation = 2.21), with schools using a range from zero to eight interventions. The average quality rating of the interventions was .42, ranging from zero to .72, indicating schools are typically using interventions with moderate empirical evidence and not those with the strongest support. The mean rating of Tier 3 evidence-based reading interventions was 1.11 (standard deviation = 0.99). Similar to Tier 2, schools typically had multiple interventions available (approximately three), but the underlying empirical support of those interventions was moderate at best. At Tier 3, the most commonly used

interventions were Tutoring (n=40), Guided Reading (n=38), and Early Intervention in Reading (n=32).

When comparing the quality of evidence between Tier 2 and Tier 3 interventions, there was a statistically significant difference in overall quality across tiers based on the composite indicator ($t = 2.542$, $p = .013$); specifically, the overall quality of interventions at Tier 2 appeared greater than the quality at Tier 3. However, in interpreting this composite, the average quality rating across tiers was statistically similar ($t = -1.024$, $p = .309$), and the greater quality at Tier 2 was actually reflective of the greater number of available interventions at Tier 2 relative to Tier 3 ($t = 4.891$, $p < .001$).

Table 6

Ratings for Reading Interventions by Database

	WWC	Promising Practices	Johns Hopkins	IES rating	Overall Code
Accelerated Reader	0	1	NR	NA	.5
Compass Learning Odyssey	NR	NR	NR	0	0
Corrective Reading	0	NR	1	NA	.5
Early Intervention in Reading	1	1	0	NA	.67
Leveled Literacy Intervention (Fountas & Pinnell)	NR	NR	NR	0	0
Orton-Gillingham	NR	NR	0	NA	0
Read 180	1	NR	1	NA	1
Read Naturally	0	NR	0	NA	0
Readers' Theatre	NR	NR	0	NA	0
Reading Mastery	0	NR	NR	NA	0
Reading Recovery	1	1	1	NA	1
Reciprocal Teaching	NR	1	0	NA	.5
SOAR to Success (Houghton-Miffin)	NR	NR	NR	0	0
Wilson Reading System	0	NR	NR	NA	0
Other					
Voyager	0	NR	0	NA	0
Imagine Learning	NR	NR	NR	0	0
Lexia	1	NR	0	NA	.5
PALS	0	1	NR	NA	.5
Sound Partners	1	NR	NR	NA	1
Early reading empowerment	NR	NR	NR	0	0
General practices					
Drill/Flashcards	NA	NA	NA	NA	.5
Guided Reading	NA	NA	NA	NA	.5
Individualized Instruction/Tutoring	NA	NA	NA	NA	.5
Partner Reading	NA	NA	NA	NA	.5
Peer Tutoring	NA	NA	NA	NA	.5

Note: NR = not-rated; NA = not applicable

Progress Monitoring. Progress monitoring tools were coded on a 0-1 point scale based on review by the National Center for Progress Monitoring, which specified the following evidence-levels: data unavailable or unconvincing evidence (0), partially convincing evidence (.5), and convincing evidence (1). If a measure had been reviewed by the National Center for Progress Monitoring, this indicated that the psychometric properties of the measure have been empirically examined. However, sub-skill mastery measures could be used for progress monitoring purposes but are not empirically validated by an external source, and therefore, were coded as partially convincing (.5). Response options coded as a 0 were those that were intended to be used as a screening measure, or that were part of an intervention package and did not necessarily represent a validated progress monitoring tool (e.g. Imagine Learning, Compass Learning, and Read 180).

The following measures were considered to have strong evidence (1): AIMSWEB, CBM-R, DIBELS, EasyCBM, STAR, and SRI, which means their psychometric properties have been empirically examined. Measures considered to have partial evidence (.5) were as follows: DRA, Running Records, district developed probes, MAP; while these measures can be used for progress monitoring, they lacked empirical validation for said purpose. Lastly the subsequent measures were rated as unconvincing: Basal Reading Series, Reading Recovery, Benchmark, Fountas & Pinnel, Compass Learning, Imagine Learning, PALS, Journeys, LLI, SRA, and Oasys; they represented an intervention package or were not intended to be used as a progress monitoring tool.

In addition to the use of a psychometrically sound progress monitoring tools, the frequency of use was an important consideration of quality because frequency impacts

the reliability and validity of the slopes of student progress, which are used for decision-making purposes. Accordingly, to create an overall composite, the frequency of progress monitoring was multiplied by the rating of the quality of the progress monitoring tool, which created a continuous indicator of quality progress monitoring.

Across the sample, the mean rating for quality of progress monitoring tools used for Tier 2 was 0.65 (standard deviation = 1.03), while the mean rating for Tier 3 progress monitoring tools was 0.77 (standard deviation = 1.22). In interpreting this finding, results indicate that at Tier 2, the average quality rating across schools of the progress monitoring tools was .68 with schools monitoring progress less than once per week. At Tier 3, the quality rating of the progress monitoring tool was .64 with schools monitoring progress 1.20 times per week. In comparing the results of Tier 2 to Tier 3, similar ratings of progress monitoring quality are noted (t = -.756, p = .452). Although trending in the appropriate direction (greater quality at Tier 3) the statistically similar quality of progress monitoring when comparing Tier 2 and Tier 3 is concerning given that decision-making in Tier 3 is typically higher-stakes (potential special education eligibility determination decisions), and it is of utmost importance to have higher psychometric quality. In addition to quality progress monitoring, treatment integrity is a requisite for the interpretation of the quality of intervention.

Treatment integrity. Descriptive categories for integrity were as follows: 0 = no method for integrity measurement, 1 = indirect measurement, 2 = direct measurement without feedback, 3 = direct measurement with feedback, 4 = direct measurement without feedback and indirect measure, 5 = direct measurement with feedback and indirect measurement. In Table 7, the frequency distribution of schools across the descriptive

categories is presented. The modal response was to have no method of integrity measurement in Tier 2 and Tier 3 (n=54 (64%) and n=47 (56%), respectively).

Table 7

Frequency Distribution of Integrity Variables

	Tier 2		Tier 3		Total	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
0	54	64.3	47	56.0	101	60.1
1	10	11.9	15	17.9	25	14.9
2	3	3.6	1	1.2	4	2.4
3	5	6.0	8	9.5	13	7.7
4	1	1.2	4	4.8	5	3.0
5	11	13.1	9	10.7	20	11.9

The mean levels of treatment integrity when comparing Tier 2 and Tier 3 were not significantly different (t = -.903, p = .369). If integrity was measured at Tier 2, the most frequently used approach was the highest quality (i.e., direct measure with feedback combined with an indirect measure). On the other hand, if integrity was measured at Tier 3, the most frequently used approach was an indirect measure only, which is a relatively lower quality approach. The final quality indicator of RtI implementation is the training of interventionists.

Interventionist training. Descriptive categories for training were as follows: 0 = no training provided, 1 = other method of training (e.g. materials presented at faculty meeting), 2 = in-service training, 3 = ongoing professional development, 4 = in-service training with ongoing professional development. In Table 8, the frequency distribution of schools across the descriptive categories is shown. At Tier 2 and Tier 3, the modal response was to have provided the highest form of training, which was in-service training with ongoing professional development (n=42 (50%) and n=30 (36%) respectively).

However, it was also common to have provided no training (n = 14 (17%) at Tier 2, and n = 24 (29%) at Tier 3).

Table 8

Frequency Distribution of Training Variables

	Tier 2		Tier 3		Total	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
0	14	16.7	24	28.6	38	22.6
1	10	11.9	5	6.0	15	8.9
2	5	5.95	6	7.1	11	6.6
3	13	15.5	19	22.6	32	19.1
4	42	50.0	30	35.7	72	42.9

Further, the average training ratings were actually worse at Tier 3 relative to Tier 2, at a level approaching statistical significance ($t = 1.766$, $p = .081$). This trend is counter to the RtI model and notion that those with the greatest expertise/training should be working with students needing the greatest support. In addition to the quality indicators of implementation, it was of interest to examine the intensity of intervention implementation across schools.

Intervention intensity. Intensity of intervention is based on a combination of intervention group size and intervention dosage. Based on the coding scheme described below, a continuous indicator of intensity was created to capture the number of minutes of intervention provided per week per student at each tier. The mean Tier 2 intervention intensity was 63.28 minutes per week per student (standard deviation = 55.58 minutes per week per student), in comparison to 113.29 minutes at Tier 3 (standard deviation = 83.28 minutes per week per student). Therefore, students in Tier 3 received nearly double the

intensity as students in Tier 2, which was a statistically significant difference ($t = 5.625$, $p < .001$).

In unpacking this finding, students at Tier 2 typically were in a group size of 2-3 (average of 2.56) and received 118 minutes per week of intervention. At Tier 3, students were in a group size of 1-2 (average of 1.40) and received approximately 135 minutes of intervention per week. Comparing across tiers, group size was significantly smaller at Tier 3 relative to Tier 2 ($t = -5.25$, $p < .001$), however, the number of minutes received was statistically similar across tiers ($t = 1.85$, $p = .068$). Overall, this was an encouraging finding given the foundational premise that Tier 3 is more intensive than Tier 2, however, this indicated that the change in intensity from Tier 3 to Tier 2 was predominantly due to a reduction in group size. In addition to providing descriptive statistics, analysis of the data for important group differences to identify potential covariates were conducted.

Covariate analysis. The correlation matrix was used to determine if there were significant relationships between various demographic variables (FRL, racial/ethnic minority, ELL, and students with disabilities), the independent variables, and the dependent variables. Table 9 presents the correlations among the demographic variables to better understand how school composition varied. Table 10 contains the correlations between the demographic variables and the independent/dependent variables to identify possible covariates and further elaborate on the relationships among these variables. If a demographic variable was found to be significantly related to a dependent variable, it was used as a predictor in the analyses for the research questions delineated below. Lastly, Table 11 presents the inter-correlations between independent and dependent variables. Specifically, significant relationships at the .01 level are described below. Given the

multitude of variables examined in the covariate analysis, the .01 level of significance was used as opposed to the .05 level of significance to ensure that chance relationships were not identified as significant covariates,

Correlations between demographic variables. Table 9 indicated significant relationships between various demographic variables. Specifically, the percent of Native American (NA) students was negatively related to the percent of White students (r = -.315, p < .01). The percent of Asian students was positively correlated to the percent of African American students (r = .539, p < .01) and Hispanic students (r = .360, p < .01), while it was negatively correlated to the percent of White students (r = -.662, p < .01) and free or reduced lunch status (r = -.357, p < .01). The percent of African American students was also positively related to the percent of Hispanic students (r = .294, p < .01) and percent of students who were English Language Learners (r = .416, p < .01); the percent of African American students was negatively related to the percent of White students (r = .659, p < .01).

Table 9

Correlation Matrix for Student Demographics

	1	2	3	4	5	6	7	8	9	10	11	12
1. NA	--											
2. Asian	.133	--										
3. Afr. Am.	.201	*.539***	--									
4. Hisp.	.077	*.360***	*.294***	--								
5. PI	.064	.143	-.040	.089	--							
6. 2 or more	.206	.140	-.107	.057	*.291***	--						
7. White	*-.315***	*-.662***	*-.659***	*-.683***	-.117	-.188	--					
8. FRL	.253*	*-.357***	.066	.037	-.128	-.053	-.071	--				
9. ELL	.091	*.589***	*.416***	*.678***	.047	.105	*-.693***	.012	--			
10. Male	-.025	-.187	-.050	-.175	-.042	-.124	.223*	-.015	-.127	--		
11. Migrant	-.061	-.049	-.196	.014	-.134	-.036	.157	-.001	-.048	.126	--	
12. Disability	.254*	-.068	-.206	.018	.175	.010	.051	.228*	.109	.075	.160	--

Note: NA = Native American; Afr. Am. = African American; Hisp = Hispanic; PI = Pacific Islander; 2 or more = 2 or more races; FRL = free/reduced lunch; ELL = English language learner

Correlations between demographic variables and independent/dependent

Variables. Table 10 indicated significant relationships between various demographic variables and the independent variables. Specifically, the percent of students scoring at minimal was positively related to free or reduced lunch status ($r = .437$, $p < .01$) and was negatively related the percent of students of Two or More races ($r = -.316$, $p < .01$). The percent of students scoring at basic was positively related to free or reduced lunch status ($r = .294$, $p < .01$). Lastly, per-pupil expenditure (PPE) was positively related to the percent of students of Two or More races ($r = .302$, $p < .01$). The only significant relationship between demographics and any of the dependent variables was the positive

association between the percent of Asian students and Tier 3 intensity ($r = .328$, $p < .01$). Accordingly, the percent of Asian students was used as a covariate for research question 4, which examined Tier 3 intensity.

Table 10

Correlation Matrix between Demographic Variables and Independent/Dependent Variables

	NA	Asian	Afr. Am.	Hisp.	PI	2 or More	White	FRL	ELL	Male	Mig.	Dis.
Minimal	-.001	-.146	.090	.142	-.140	*-.316***	-.076	*.437***	.234*	-.043	.150	.210
Basic	.116	-.252*	-.190	-.124	-.005	.093	.179	*.294***	-.083	.150	.161	.084
PPE	-.052	-.272*	-.201	.105	.097	*.302***	.022	.191	-.041	-.024	-.248*	-.033
T2EBI	.093	.062	.030	.084	-.003	.066	-.145	-.048	.126	.004	.043	.086
T2PM	.135	-.104	-.102	.072	-.073	.070	-.010	-.045	-.022	-.089	.107	-.033
T2Training	.095	.069	.147	.049	-.077	-.076	-.171	-.032	.207	.060	-.003	.090
T2Integrity	-.060	-.143	.001	-.140	-.194	-.087	.088	-.083	-.058	.228*	.114	-.161
T2Intensity	.077	.107	.145	.103	-.018	-.062	-.229*	-.061	.182	.002	-.051	.116
T3EBI	-.109	-.031	-.144	-.030	-.156	.014	.018	-.111	.078	.058	.282*	.084
T3PM	-.080	.132	.031	-.009	.017	.111	-.078	-.243*	.091	.000	.152	.044
T3Training	-.042	.004	.094	-.143	-.193	-.092	.000	-.003	.051	.105	.086	-.088
T3Integrity	-.030	-.131	-.078	-.254*	-.124	.004	.153	-.078	-.171	.110	.074	-.113
T3Intensity	-.114	*.328***	.189	-.050	-.228*	-.065	-.162	-.236*	.155	.202	.051	-.076

Note: NA = Native American; Afr. Am. = African American; Hisp. = Hispanic; PI = Pacific Islander; 2 or more = 2 or more races; FRL = free/reduced lunch; ELL = English language learner; Mig. = Migrant; Dis. = students with disabilities.

Correlations between independent and dependent variables. Table 12 presents the inter-correlations among the independent and dependent variables. For the purpose of this study, both independent variables (achievement levels and per-pupil expenditure) were treated as continuous variables. None of the independent variables were significantly related to each other and accordingly, percent minimal and percent basic

were treated as separate independent variables. Since the percent of students at minimal was not significantly correlated to the percent of students scoring basic, these two variables may capture different profiles of achievement. Additionally, per-pupil expenditure was not significantly related to the percent of students scoring minimal or the percent of students scoring basic.

There were no significant relationships between independent variables and dependent variables, but there were many significant relationships between the dependent variables. Specifically, Tier 2 EBI was positively related to Tier 2 PM ($r = .437$, $p < .01$), Tier 2 Training ($r = .489$, $p < .01$), Tier 2 Intensity ($r = .573$, $p < .01$), Tier 3 EBI ($r = .432$, $p < .01$) and Tier 3 PM ($r = .335$, $p < .01$). Tier 2 PM was positively related to Tier 2 Training ($r = .396$, $p < .01$), Tier 2 Integrity ($r = .363$, $p < .01$), Tier 2 Intensity ($r = .485$, $p < .01$), and Tier 3 PM ($r = .459$, $p < .01$). Tier 2 Training was positively related to Tier 2 Integrity ($r = .361$, $p < .01$), Tier 2 Intensity ($r = .581$, $p < .01$), Tier 3 PM ($r = .322$, $p < .01$), Tier 3 Training ($r = .416$, $p < .01$), and Tier 3 Intensity ($r = .341$, $p < .01$). Tier 2 Integrity was positively related to Tier 3 PM ($r = .288$, $p < .01$), Tier 3 Training ($r = .335$, $p < .01$), Tier 3 Integrity ($r = .451$, $p < .01$), and Tier 3 Intensity ($r = .308$, $p < .01$). Tier 2 Intensity was positively related to Tier 3 Intensity ($r = .372$, $p < .01$).

Tier 3 EBI was positively related to Tier 3 PM ($r = .448$, $p < .01$), Tier 3 Training ($r = .430$, $p < .01$), and Tier 3 Intensity ($r = .337$, $p < .01$). Tier 3 PM was positively related to Tier 3 Training ($r = .382$, $p < .01$), Tier 3 Integrity ($r = .420$, $p < .01$), and Tier 3 Intensity ($r = .449$, $p < .01$). Lastly, Tier 3 Training was positively related to Tier 3 Integrity ($r = .374$, $p < .01$) and Tier 3 Intensity ($r = .490$, $p < .01$). The inter-relationships between dependent variables suggests that certain schools tend to have

higher overall levels of implementation across many indicators, while others were low across many indicators. For example, those higher on one indicator of quality tended to be higher on other indicators of quality and vice versa. This suggests RtI implementation may not necessarily be piece-meal and implementation may develop relatively uniformly across components within a school (or in stages).

Table 11

Correlation Matrix for Independent and Dependent Variables

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Minimal	--												
2. Basic	.046	--											
3. PPE	-.195	.047	--										
4. T2EBI	.003	-.158	-.089	--									
5. T2PM	-.077	.094	.125	*.437***	--								
6. T2Training	.135	-.039	-.074	*.489***	*.396***	--							
7. T2Integrity	-.059	.146	-.104	.249*	*.363***	*.361***	--						
8. T2Intensity	.144	-.179	-.009	*.573***	*.485***	*.581***	.140	--					
9. T3EBI	.048	-.044	-.059	*.432***	.256*	.255*	.198	.191	--				
10. T3PM	-.188	.006	.067	*.335***	*.459***	*.322***	*.288***	.233*	*.448***	--			
11. T3Training	.048	.046	-.079	.218	.148	*.416***	*.335***	.229*	*.430***	*.382***	--		
12. T3Integrity	-.157	.122	-.096	.039	.245*	.095	*.451***	-.069	.251*	*.420***	*.374***	--	
13. T3Intensity	-.186	-.106	-.069	.227*	.277*	*.341***	*.308***	*.372***	*.337***	*.449***	*.490***	.223*	--

Summary. Collectively, descriptive analyses indicated that there was a statistically significant difference in overall evidence-based intervention quality across tiers based on the composite indicator; specifically, the overall quality of interventions at Tier 2 was greater than the quality at Tier 3, which was due to the greater number of available interventions. However, statistically similar results were found when comparing

tiers across the remaining quality indicators (progress monitoring, treatment integrity, and training). Lastly, when comparing intervention intensity across tiers, students in Tier 3 received nearly double the intensity as students in Tier 2, which was a statistically significant difference. This is an encouraging finding given the foundational premise that Tier 3 is more intensive than Tier 2.

Many of the demographic variables were significantly correlated with each other. Specifically, the percent of students of various minority races tended to be positively related to other minority races (e.g. schools with higher African American populations often had higher populations of Hispanic students), and they were negatively related to the proportion of White students. In examining the relationships between demographics and independent/dependent variables, there were significant correlations between free or reduced lunch status and the percent of students scoring minimal and basic; this is consistent with many findings demonstrating the relationship between poverty and low academic achievement (Equity and Excellence Commission, 2013). Additionally, the only significant relationship between demographics and any of the dependent variables was the positive association between the percent of Asian students and Tier 3 intensity, which was therefore used as a covariate for research question 4. No significant relationships were observed between the independent variables, while many significant positive relationships were observed between dependent variables. The inter-relationships between dependent variables suggests that certain schools tend to have higher overall levels of implementation across indicators while others are low across indicators. Collectively, these findings frame the examination of the relationship

between per-pupil expenditure, achievement levels, and the implementation of quality reading interventions.

Regression Analyses

Assumptions for Regression Analysis. The following assumptions were examined in the context of the univariate regression analyses: linearity, homoscedasticity, normality, and multicollinearity. Additionally, outliers were examined using Mahalanobis distances.

Assumptions. Examination of residual plots indicated that assumptions of linearity and homoscedasticity were not violated. Tests of normality indicated violations with Asian, Hispanic, English Language Learners (Limited English Proficient), and per-pupil expenditure (based on Shapiro-Wilk test). Log(10) transformations were used for all four variables and served to create appropriately normal distributions. Multicollinearity was examined with Tolerance levels; all variables were above .1, indicating that multicollinearity levels were not severe.

Outliers. Outliers were examined using Mahalanobis distances. Six outliers were detected and removed based on a critical value of 18.5, with df = 4 (number of variables including covariates) and $p < .001$. Three schools were outliers based on the percent of students performing at the minimal level, two schools were outliers based on per-pupil expenditure, and one school was an outlier based on the percent of students performing at the basic level. All outliers were removed prior to the regression analyses.

Question 1: Tier 2 Reading Interventions by PPE and Reading Proficiency

The primary research question of interest was whether the use of high-quality reading interventions (i.e., the quality of evidence-based interventions, progress

monitoring tools, treatment integrity measures, and use of trained interventionists) was predicted by per-pupil expenditure and students' reading proficiency levels at Tier 2. More specifically, the following research questions were addressed in this analysis.

1a. Does per-pupil expenditure and student's reading proficiency predict the use of quality of evidence-based interventions at Tier 2? In testing the relationship between reading proficiency, per-pupil expenditure and evidence-based Tier 2 interventions; reading proficiency (percent minimal and percent basic) was entered as block 1, and per-pupil expenditure was entered as block 2 (see Tables 12-13). Model 1, with percent minimal and basic as the sole predictors, was not significantly predictive of the Tier 2 evidence-based interventions, and explained 4.0% of the variance in the outcome variable, $R^2 = .040$, $F(2, 75) = 1.572$, $p = .214$. Model 2, with all three independent variables was also not significantly predictive of Tier 2 evidence-based interventions and explained 4.9% of the variance in the outcome variable, $R^2 = .049$, $F(3, 74) = 1.258$, $p = .295$. However, the decrease in the adjusted R^2 from model 1 to model 2 indicated that the increase in predictive power can be explained by chance alone. This was further reiterated by the non-significant F change from model 1 to model 2, $F(1, 74) = .645$, $p = .425$. Collectively, this suggested that reading proficiency alone, and the combination of reading proficiency and per-pupil expenditure did not predict the quality of evidence-based intervention implementation at Tier 2.

Table 12

Model Summary for Research Question 1a

					Change Statistics				
Model	*R*	R^2	*Adjusted* R^2	*SE of the Estimate*	R^2 *Change*	*F Change*	*df1*	*df2*	*Sig. F Change*
1	.201[a]	.040	.015	.981	.040	1.572	2	75	.214
2	.220[b]	.049	.010	.983	.008	.645	1	74	.425

a. Predictors: (Constant), Basic, Minimal

b. Predictors: (Constant), Basic, Minimal, PPE

Table 13

ANOVA for Research Question 1a

Model		*SS*	*df*	*MS*	*F*	*Sig.*
1	Regression	3.026	2	1.513	1.572	.214[a]
	Residual	72.184	75	.962		
	Total	75.21	77			
2	Regression	3.649	3	1.216	1.258	.295[b]
	Residual	71.561	74	.967		
	Total	75.21	77			

a. Predictors: (Constant), Basic, Minimal

b. Predictors: (Constant), Basic, Minimal, PPE

c. Dependent Variable: T2EBI

In Table 14, the coefficients and significance of the predictors in each model is described. The β coefficients indicated a negative relationship between all predictors (minimal, basic, and PPE and the outcome variable) in both models. However, none of the relationships were statistically significant (in both models).

Table 14

Summary of Regression Model for Research Question 1a

Model		*B*	*SE*	*β*	*t*	*Sig.*
1	(Constant)	3.023	.944			
	Minimal	-.232	1.654	-.016	-.140	.889
	Basic	-3.96	2.236	-.200	-1.771	.081
2	(Constant)	4.207	1.752			
	Minimal	-.484	1.687	-.033	-.287	.775
	Basic	-3.951	2.241	-.200	-1.763	.082
	PPE	$-9.22E^{-05}$	.000	-.093	-.803	.425

a. Dependent Variable: T2EBI

1b. Does per-pupil expenditure and student's reading proficiency predict the use of quality progress monitoring tools at Tier 2? As previously noted, the relationship between reading proficiency, per-pupil expenditure and the use of quality progress monitoring tools at Tier 2 was examined. Reading proficiency (percent minimal and percent basic) was entered as block 1, and per-pupil expenditure was entered as block 2. Model 1 (minimal and basic) was not significantly predictive of Tier 2 progress monitoring, and explained 4.9% of the variance in the outcome variable $R^2 = .049$, $F(2, 75) = 1.939$, $p = .151$. The model with all predictors (Minimal, Basic, and PPE) predicted a significant portion of variance in the quality of Tier 2 progress monitoring; specifically, the model explained 10.1% of the variance, $R^2 = .101$, $F(2, 75) = 2.767$, $p = .043$. The increase in adjusted R^2 from model 1 to model 2 corroborates this finding as there was a statistically significant F change, $F(1, 74) = 4.253$, $p = .043$. Collectively, this indicated that reading proficiency does not significantly predict the quality of progress monitoring at Tier 2, but reading proficiency in concert with per-pupil expenditure did

significantly impact the quality of progress monitoring at Tier 2. However, the model was not significantly predictive of the outcome variable when restricting the family-wise alpha to .05 and thereby making the necessary p-value for significance equal to .05/4 = .0125.

Table 15

Model Summary for Research Question 1b

					Change Statistics				
Model	*R*	R^2	*Adjusted* R^2	*SE of the Estimate*	R^2 *Change*	*F Change*	*df1*	*df2*	*Sig. F Change*
1	.222[a]	.049	.024	1.032	.049	1.989	2	75	.151
2	.318[b]	.101	.064	1.010	.052	4.253	1	74	.043

a. Predictors: (Constant), Basic, Minimal

b. Predictors: (Constant), Basic, Minimal, PPE

Table 16

ANOVA for Research Question 1b

Model		*SS*	*df*	*MS*	*F*	Sig.
1	Regression	4.135	2	2.068	1.939	.151[a]
	Residual	79.964	75	1.066		
	Total	84.100	77			
2	Regression	8.841	3	2.827	2.767	.048[b]
	Residual	75.618	74	1.022		
	Total	84.100	77			

a. Predictors: (Constant), Basic, Minimal

b. Predictors: (Constant), Basic, Minimal, PPE

c. Dependent Variable: T2PM

In Table 17, coefficients and significance of the predictors in each model are delineated. Neither percent minimal nor percent basic was significantly predictive of the

outcome variable in either model, however, per-pupil expenditure was significantly predictive of the quality of Tier 2 progress monitoring, $t = 2.062$, $p = .043$. In interpreting this finding, the positive directionality of β (.231) indicates that on a standardized metric, an increase in per-pupil expenditure is significantly related to an increase in quality of progress monitoring at Tier 2. Similar to above, the PPE predictor was not significantly predictive of the outcome variable when restricting the family-wise alpha to .05 and thereby making the necessary p-value for significance equal to .05/4 = .0125.

Table 17

Summary of Regression Model for Research Question 1b

Model		*B*	*SE*	*β*	*t*	Sig.
1	(Constant)	-.766	.994			
	Minimal	-.806	1.741	-.052	-.463	.645
	Basic	4.465	2.353	.214	1.898	.062
2	(Constant)	-3.891	1.801			
	Minimal	-.142	1.734	-.009	-.082	.935
	Basic	4.442	2.304	.213	1.928	.058
	PPE	.000	.000	.231	2.062	.043

a. Dependent Variable: T2PM

1c. Does per-pupil expenditure and student's reading proficiency predict the use of quality treatment integrity measures at Tier 2? Due to the ordinal nature of outcome variable, ordinal regression was used to address the relationship between reading proficiency, per-pupil expenditure and their effect on Tier 2 treatment integrity. The Chi-Square statistic indicated that the model did not provide a significant improvement over the baseline intercept-only model and explained 3.7% of the variance in the outcome

variable, $R^2 = .037$, $\chi^2(3) = 2.650$, $p = .449$. This result signifies that reading proficiency and per-pupil expenditure do not predict the quality of treatment integrity measures at Tier 2.

Table 18

Model Summary for Research Question 1c

Model	χ^2	*df*	*sig*	R^2 *(Nagelkerke)*
Minimal, Basic, PPE	2.650	3	.449	.037

Table 19 indicates the parameter estimates for the model with all three predictor variables (minimal, basic, and PPE). None of the predictor variables significantly impacted the odds of quality integrity measures at Tier 2.

Table 19

Summary of Parameter Estimates for Research Question 1c

		Estimate	*Std. Error*	*Wald*	*df*	*Sig.*
Threshold	[T2Integrity = 0]	-1.093	3.926	.077	1	.781
	[T2Integrity = 1]	-.467	3.926	.014	1	.905
	[T2Integrity = 2]	-.319	3.926	.007	1	.935
	[T2Integrity = 3]	.12	3.928	.001	1	.976
	[T2Integrity = 4]	.225	3.929	.003	1	.954
Location	Minimal	-.289	3.596	.006	1	.936
	Basic	5.506	4.558	1.459	1	.227
	PPE	.000	.000	1.193	1	.275

Link function: Logit.

1d. Does per-pupil expenditure and student's reading proficiency predict the use of quality interventionist training at Tier 2? In examining the relationship between reading proficiency, per-pupil expenditure and their effect on quality training at Tier 2,

ordinal regression was used. The Chi-Square statistic indicated that the model did not provide a significant improvement over the baseline intercept-only model and explained 4.3% of the variance in the outcome variable, $R^2 = .043$, $\chi^2(3) = 3.169$, $p = .366$. This result signifies that reading proficiency and per-pupil expenditure do not predict the quality of interventionist training at Tier 2.

Table 20

Model Summary for Research Question 1d

Model	*Chi-square*	*df*	*sig*	*R^2 (Nagelkerke)*
Minimal, Basic, PPE	3.169	3	.366	.043

Table 21 indicates the parameter estimates for the model with all three predictor variables (Minimal, Basic, and PPE). None of the predictor variables significantly impacted the odds of quality interventionist training at Tier 2.

Table 21

Summary of Parameter Estimates for Research Question 1d

		Estimate	*Std. Error*	*Wald*	*df*	*Sig.*
Threshold	[T2Training = 0]	-4.831	3.309	2.131	1	.144
	[T2Training = 1]	-4.049	3.296	1.509	1	.219
	[T2Training = 2]	-3.740	3.291	1.291	1	.256
	[T2Training = 3]	-3.021	3.281	.848	1	.357
Location	Minimal	3.340	3.243	1.061	1	.303
	Basic	-2.719	4.231	.413	1	.520
	PPE	.000	.000	1.076	1	.300

Link function: Logit.

Question 2: Tier 3 Reading Interventions by PPE and Reading Proficiency

The second primary research question of interest was whether the use of high-quality reading interventions (i.e., the quality of evidence-based interventions, progress monitoring tools, treatment integrity measures, and use of trained interventionists) was predicted by per-pupil expenditure and students' reading proficiency levels at Tier 3. More specifically, the following research questions were addressed in this analysis.

2a. Does per-pupil expenditure and student's reading proficiency predict the use of quality of evidence-based interventions at Tier 3? In testing the relationship between reading proficiency, per-pupil expenditure and evidence-based Tier 3 interventions; reading proficiency (percent minimal and percent basic) was entered as block 1, and per-pupil expenditure was entered as block 2 (see Tables 22-23). Model 1, with percent minimal and basic as the sole predictors, was not significantly predictive of Tier 2 evidence-based interventions, and explained 0.7% of the variance in the outcome variable, $R^2 = .007$, $F(2, 75) = .281$, $p = .756$. Model 2, with all three independent variables was also not significantly predictive of Tier 2 evidence-based interventions and explained 1.1% of the variance in the outcome variable, $R^2 = .011$, $F(3, 74) = .270$, $p = .847$. However, the decrease in the adjusted R^2 from model 1 to model 2 indicated that the increase in predictive power can be explained by chance alone. This was further reiterated by the non-significant F change from model 1 to model 2, $F(1, 74) = .281$, $p = .616$. Collectively, this suggested that reading proficiency, and the combination of reading proficiency and per-pupil expenditure did not predict the quality of evidence-based intervention implementation at Tier 3.

Table 22

Model Summary for Research Question 2a

					Change Statistics				
Model	*R*	R^2	*Adjusted* R^2	*SE of the Estimate*	R^2 *Change*	*F Change*	*df1*	*df2*	*Sig. F Change*
1	.086[a]	.007	-.019	.998	.007	.281	2	75	.756
2	.104[b]	.011	-.029	1.002	.003	.253	1	74	.616

a. Predictors: (Constant), Basic, Minimal

b. Predictors: (Constant), Basic, Minimal, PPE

Table 23

ANOVA for Research Question 2a

Model		*SS*	*df*	*MS*	*F*	*Sig.*
1	Regression	.560	2	.280	.281	.756[a]
	Residual	74.683	75	.996		
	Total	75.242	77			
2	Regression	.815	3	.272	.270	.847[b]
	Residual	74.428	74	1.006		
	Total	75.242	77			

a. Predictors: (Constant), Basic, Minimal

b. Predictors: (Constant), Basic, Minimal, PPE

c. Dependent Variable: T3EBI

Table 24 describes coefficients and significance of the predictors in each model. As is evidenced in the table, the β coefficients indicated a negative relationship between all predictors (minimal, basic, and PPE) in both models and the outcome variable. However, none of the relationships were statistically significant (in both models).

Table 24

Summary of Regression Model for Research Question 2a

Model		*B*	*SE*	*β*	*t*	*Sig.*
1	(Constant)	1.795	.961			
	Minimal	-.313	1.682	-.021	-.186	.853
	Basic	-1.665	2.274	-.084	-.732	.466
2	(Constant)	2.552	1.787			
	Minimal	-.474	1.721	-.032	-.276	.784
	Basic	-1.659	2.285	-.084	-.726	.470
	PPE	$-5.89E^{-5}$	.000	-.059	-.503	.616

a. Dependent Variable: T3EBI

2b. Does per-pupil expenditure and student's reading proficiency predict the use of quality progress monitoring tools at Tier 3? As previously noted, the relationship between reading proficiency, per-pupil expenditure and the use of quality progress monitoring tools at Tier 3 was examined. Reading proficiency (percent minimal and percent basic) was entered as block 1, and per-pupil expenditure was entered as block 2. Model 1 (minimal and basic) was not significantly predictive of Tier 3 progress monitoring, and explained 1.6% of the variance in the outcome variable, $R^2 = .016$, $F(2, 75) = .614$, $p = .544$. The model with all predictors (minimal, basic, and PPE) also did not predict a significant portion of variance in the quality of Tier 3 progress monitoring; specifically, the model explained 2.1% of the variance, $R^2 = .021$, $F(2, 75) = .536$, $p = .659$. The decrease in the adjusted R^2 from model 1 to model 2 indicated that the increase in predictive power can be explained by chance alone. This was further reiterated by the non-significant F change from model 1 to model 2, $F(1, 74) = .389$, $p = .535$. Collectively, this suggested that reading proficiency, and the combination of reading

proficiency and per-pupil expenditure did not predict the quality of progress monitoring at Tier 3.

Table 25

Model Summary for Research Question 2b

					Change Statistics				
Model	*R*	R^2	*Adjusted* R^2	*SE of the Estimate*	R^2 *Change*	*F Change*	*df1*	*df2*	*Sig. F Change*
1	.127[a]	.016	-.010	1.261	.016	.614	2	75	.544
2	.146[b]	.021	-.018	1.267	.005	.389	1	74	.535

a. Predictors: (Constant), Basic, Minimal

b. Predictors: (Constant), Basic, Minimal, PPE

Table 26

ANOVA for Research Question 2b

Model		*SS*	*df*	*MS*	*F*	*Sig.*
1	Regression	1.956	2	.978	.614	.544[a]
	Residual	119.422	75	1.592		
	Total	121.378	77			
2	Regression	2.581	3	.860	.536	.659[b]
	Residual	118.797	74	1.605		
	Total	121.378	77			

a. Predictors: (Constant), Basic, Minimal

b. Predictors: (Constant), Basic, Minimal, PPE

c. Dependent Variable: T3PM

Table 27 delineates coefficients and significance of the predictors in each model. As is evidenced in the table, none of the predictor variables (minimal, basic, and PPE) in both models were significantly related to the outcome variable.

Table 27

Summary of Regression Model for Research Question 2b

Model		*B*	*SE*	*β*	*t*	*Sig.*
1	(Constant)	-.224	1.215			
	Minimal	-.334	2.217	-.018	-.157	.876
	Basic	3.138	2.875	.125	1.091	.279
2	(Constant)	.961	2.257			
	Minimal	-.586	2.174	-.032	-.270	.788
	Basic	3.147	2.887	.125	1.090	.279
	PPE	$-9.23E^{-5}$	.000	-.073	-.624	.535

a. Dependent Variable: T2PM

2c. Does per-pupil expenditure and student's reading proficiency predict the use of quality treatment integrity measures at Tier 3? Due to the ordinal nature of outcome variable, ordinal regression was used to address the relationship between reading proficiency, per-pupil expenditure and their effect on Tier 3 treatment integrity. The Chi-Square statistic indicated that the model did not provide a significant improvement over the baseline intercept-only model and explained 4.6% of the variance in the outcome variable, $R^2 = .046$, $\chi^2(3) = 3.391$, $p = .335$. This result signifies that reading proficiency and per-pupil expenditure do not predict the quality of treatment integrity measures at Tier 3.

Table 28

Model Summary for Research Question 2c

Model	*Chi-square*	*df*	*sig*	*R^2 (Nagelkerke)*
Minimal, Basic, PPE	3.391	3	.335	.046

Table 29 indicates the parameter estimates for the model with all three predictor variables (minimal, basic, and PPE). None of the predictor variables significantly impacted the odds of quality integrity measures at Tier 3. Note that an Estimate is not provided for T3Integrity = 2, because after the removal of outliers, this response option was not selected within the sample.

Table 29

Summary of Parameter Estimates for Research Question 2c

		Estimate	*Std. Error*	*Wald*	*df*	*Sig.*
Threshold	[T2Integrity = 0]	-.274	3.610	.006	1	.939
	[T2Integrity = 1]	.606	3.612	.028	1	.867
	[T2Integrity = 3]	1.259	3.617	.121	1	.728
	[T2Integrity = 4]	1.738	3.623	.230	1	.631
Location	Minimal	-2.222	3.421	.422	1	.516
	Basic	6.678	4.387	2.317	1	.128
	PPE	.000	.000	.722	1	.396

Link function: Logit.

2d. Does per-pupil expenditure and student's reading proficiency predict the use of quality interventionist training at Tier 3? In examining the relationship between reading proficiency, per-pupil expenditure and their effect on quality training at Tier 3, ordinal regression was used. The Chi-Square statistic indicated that the model did not provide a significant improvement over the baseline intercept-only model and explained .9% of the variance in the outcome variable, $R^2 = .009$, $\chi^2(3) = .645$, $p = .886$. This result signifies that reading proficiency and per-pupil expenditure do not predict the quality of interventionist training at Tier 3.

Table 30

Model Summary for Research Question 2d

Model	*Chi-square*	*df*	*sig*	*R^2 (Nagelkerke)*
Minimal, Basic, PPE	.645	3	.886	.009

Table 31 indicates the parameter estimates for the model with all three predictor variables (minimal, basic, and PPE). None of the predictor variables significantly impacted the odds of quality interventionist training at Tier 3.

Table 31

Summary of Parameter Estimates for Research Question 2d

		Estimate	*Std. Error*	*Wald*	*df*	*Sig.*
Threshold	[T2Training = 0]	-2.009	3.222	.389	1	.533
	[T2Training = 1]	-1.716	3.220	.284	1	.594
	[T2Training = 2]	-1.390	3.218	.186	1	.666
	[T2Training = 3]	-.426	3.214	.018	1	.895
Location	Minimal	1.573	3.101	.257	1	.612
	Basic	-.360	4.109	.008	1	.930
	PPE	.000	.000	.256	1	.613

Link function: Logit.

Question 3: Tier 2 Intensity by Reading Proficiency

3. Does student's reading proficiency predict the intensity (based on dosage and group size) of intervention implementation at Tier 2? In testing the relationship between reading proficiency and Tier 2 intervention intensity, minimal and basic were used as predictors to examine their effect on intervention intensity. The model was not significantly predictive of Tier 2 intensity, and explained 5% of the variance in the

outcome variable, $R^2 = .050$, $F(2, 75) = 1.990$, $p = .144$. Collectively, this suggested that reading proficiency was not predictive of intervention intensity at Tier 2.

Table 32

Model Summary for Research Question 3

					Change Statistics				
Model	*R*	R^2	*Adjusted* R^2	*SE of the Estimate*	R^2 *Change*	*F Change*	*df1*	*df2*	*Sig. F Change*
1	.224[a]	.050	.025	55.058	.050	1.990	2	75	.144

a. Predictors: (Constant), Basic, Minimal

Table 33

ANOVA for Research Question 3

Model		*SS*	*df*	*MS*	*F*	*Sig.*
1	Regression	12064.675	2	6032.337	1.990	.144[a]
	Residual	227356.337	75	3031.418		
	Total	239421.012	77			

a. Predictors: (Constant), Basic, Minimal

c. Dependent Variable: T2Intensity

Table 34 delineates coefficients and significance of the predictors in the model. As is evidenced in the table, none of the predictor variables (minimal and basic) were significantly related to the outcome variable.

Table 34

Summary of Regression Model for Research Question 3

Model		*B*	*SE*	*β*	*t*	*Sig.*
1	(Constant)	-.224	1.215			
	Minimal	-.334	2.217	-.018	-.157	.876
	Basic	3.138	2.875	.125	1.091	.279

a. Dependent Variable: T2Intensity

Question 4: Tier 3 Intensity by Reading Proficiency

4. Does student's reading proficiency predict the intensity (based on dosage and group size) of intervention implementation at Tier 3? Because the percent of Asian students was significantly correlated ($p<.01$) to the outcome variable (Tier 3 intensity), it was entered as a covariate. Therefore, Asian was entered as block 1, while minimal and basic were entered as block 2 in examining their effect on Tier 3 intervention intensity. Model 1 (Asian) was significantly predictive of Tier 3 intervention intensity, and explained 15.5% of the variance in the outcome variable $R^2 = .155$, $F(1, 76) = 11.374$, $p < .001$. The model with all predictors (Asian, minimal, basic) also predicted a significant portion of variance in the quality of Tier 3 intervention intensity; specifically, the model explained 18.8% of the variance, $R^2 = .188$, $F(3, 74) = 4.635$, $p = .006$. The decrease in the adjusted R^2 from model 1 to model 2 indicated that the increase in predictive power can be explained by chance alone. This was further reiterated by the non-significant F change from model 1 to model 2, $F(2, 74) = 1.224$, $p = .301$.

Table 35

Model summary for Research Question 4

					Change Statistics				
Model	*R*	R^2	*Adjusted* R^2	*SE of the Estimate*	R^2 *Change*	*F Change*	*df1*	*df2*	*Sig. F Change*
1	.394[a]	.155	.141	74.652	.155	11.374	1	76	.001
2	.434[b]	.188	.148	74.383	.033	1.224	2	74	.301

a. Predictors: (Constant), Asian

b. Predictors: (Constant), Asian, Basic, Minimal

Table 36

ANOVA for Research Question 4

Model		*SS*	*df*	*MS*	*F*	*Sig.*
1	Regression	63387.826	1	63387.826	11.374	.001[a]
	Residual	345518.326	76	5572.876		
	Total	408906.152	77			
2	Regression	76935.641	3	25645.214	4.635	.006[b]
	Residual	331970.512	74	5532.842		
	Total	408906.152	77			

a. Predictors: (Constant), Asian

b. Predictors: (Constant), Asian Basic, Minimal

c. Dependent Variable: T3Intensity

Table 37 delineates coefficients and significance of the predictors in each model. In model 1, the percent of Asian students was significantly predictive of intervention intensity at Tier 3, $t = 3.373$, $p = .001$. In interpreting this finding, the positive directionality of β (.394) indicates that, on a standardized metric, an increase in the percent of Asian students is significantly related to an increase in intervention intensity Tier 3. In examining model 2, the percent of Asian students remained a significant

predictor of intervention intensity at Tier 3, however, neither percent minimal nor percent basic were significantly predictive of the outcome variable.

Table 37

Summary of Regression Model for Research Question 4

Model		B	SE	β	t	Sig.
1	(Constant)	98.963	10.530			
	Asian	58.332	17.296	.394	3.373	.001
2	(Constant)	222.301	85.589			
	Asian	52.614	17,645	.355	2.982	.004
	Minimal	-144.963	134.627	-.126	-1.077	.286
	Basic	-240.287	204.929	-.139	-1.173	.246

a. Dependent Variable: T3Intensity

CHAPTER FIVE

Discussion

The purpose of this study was to examine how financial capacity and achievement levels contribute to the implementation of high-quality reading interventions in the context of RtI. Specifically, the following four components were used as indicators of high-quality reading interventions in the context of RtI at both Tier 2 and Tier 3: evidence-based interventions, progress monitoring tools, treatment integrity measures, and trained interventionists. As a secondary research interest, the relationship between reading proficiency and intervention intensity (based on dosage and group size) was examined at both Tier 2 and Tier 3.

The Role of Achievement Levels and Per-pupil Expenditure in RtI

Response-to-intervention has been proposed as a movement that can potentially address the challenges of students at risk for reading failure, by promoting systems of preventative, evidence-based interventions (American Psychological Association, 2005; National Association of School Psychologists, 2007; National Joint Committee on Learning Disabilities, 2005). The implementation of interventions is a substantial requirement for schools, and yet surprisingly, the implementation of interventions has commonly been assumed rather than considered a major issue to be resolved (Gansle & Noell, 2007). Understanding the barriers and facilitators to implementation are critically important in the context of RtI because the model relies on the uptake of the empirically-based interventions in the context of routine practice.

Despite the importance of implementation, high-quality evidence-based reading interventions can be costly in terms of money and staff. However, state funding systems

rely, largely, on local property taxes, which enables districts in high-wealth parts of a state to be funded more generously than districts in low-wealth areas (Epstein, 2011). The faulty system of school funding highlights a potentially glaring challenge for the provision of high-quality reading interventions in schools. Evidence-based reading interventions are part of larger legislative aims (NCLB and IDEA), which target students at risk for difficulties, and yet, schools in lower-income areas receive less funding than those in wealthier areas due to the dependence on local funding sources. This engenders a contradiction where schools with the greatest need for the provision of high-quality evidence-based reading interventions may not be provided with the financial support to do so.

In addition to financial resources, the achievement levels of a student population might impact school personnel's capacity to implement high-quality reading interventions. The impact may occur indirectly and directly. The indirect impact of achievement can take place by creating a greater financial strain on schools. Specifically, if achievement levels do not fit the 80-15-5 template theorized by RtI, then low achieving schools will require greater amounts of each intervention implementation component. In this manner, low achievement may create additional financial strain on a school, which as described above, is hypothesized to contribute to the ability to implement high-quality reading interventions. Achievement may also directly impact the ability to implement interventions in adherence to the recommended practices of intervention implementation. Specifically, low achieving schools will have more students in need of tiered supports (i.e. intervention and progress monitoring). Having large proportions of low achieving students may create logistical challenges in providing interventions at the recommended

dosage and group size, in other words, at the recommended intensity. Based on the above rationale, the purpose of this study was to examine how financial capacity and achievement levels contribute to the implementation of high-quality reading interventions in the context of RtI.

The present findings suggest that reading proficiency and per-pupil-expenditure are not significantly predictive of the use of high-quality reading interventions at Tier 2 or Tier 3. Specifically, significant relationships were not observed between these two predictors and any of the four quality indicators: evidence-based interventions, progress monitoring tools, treatment integrity measures, and use of trained interventionists. These findings are inconsistent with the above logic hypothesizing that schools would be systematically impacted by funding and achievement levels.

The lack of significant relationships between per-pupil expenditure, achievement levels, and the quality of evidence-based reading interventions is hypothesized to have occurred for two inter-related reasons: 1) schools appear to be at varying stages in implementation, and 2) most schools appear to be at an early stage of implementation.

To specify, given the significant correlations among the dependent variables (i.e. nearly all combinations of the Tier 2 and Tier 3 dependent variables were significantly correlated at the .05 level), it is hypothesized that schools were at varying stages of RtI implementation (i.e., there are schools that are at high levels of implementation and schools that are at low levels of implementation). As conceptualized by Fixsen et al., (2005) implementation is a process not an event. The process is governed by sequential stages including: 1) exploration and adoption, 2) program installation, 3) initial implementation, 4) full implementation, 5) innovation, and 6) sustainability. Given the

legal platform for the RtI model in the state of interest, exploration and adoption was not necessarily a choice that can be made by schools, however, being that the model was not yet legally required, schools may have varied considerably in their development and preparation for the implementation of RtI. Given the relatively low average quality ratings on the indicators of quality interventions, it was likely the case that many schools were at an early stage of implementation. Specifically, quality ratings ranged from zero to less than two, with the exception of the training indicator. The higher rating on the training indicator was consistent with the notion that schools were actively preparing for the implementation of RtI by developing structural supports for the implementation of the program; active preparation and the development of structural supports are characterizations of the program installation phase of implementation (Fixsen et al., 2005), which is the second stage of implementation.

Second, based on the relatively low ratings of quality across indicators, it appeared as though many schools were at an early stage of implementation. The impact of school-level funding and school-level achievement may be too distally related to the provision of interventions at this point in implementation. For example, Jimerson, Burns, and VanDerHeyden (2007) highlight that school finances are one of many variables that may lead to successful wide-scale implementation, such as time, resources, leadership, strategic planning, preparation of professionals (beyond intervention training), empirical evidence, and evaluation of implementation. Further, this finding was actually consistent with the implementation science literature, which focuses more so on the following variables as predictors of implementation: quality of communication between the purveyor and organization, extent to which an organization has a learning culture and

central decision-making structure, and the extent to which implementation was seen as relatively easy and compatible with the organizations philosophy (Fixsen et al., 2005; Sanetti & Kratochwill, 2009). Overall implementation effectiveness was positively related to having a system in place for monitoring implementation progress, access to technical assistance, perceived ability of the organization to manage risks, and belief in the scientific evidence in support of the program (Fixsen et al., 2005). The aforementioned variables have potentially differential effects depending on the stage of implementation, which further clouds the results of the current study (e.g. it is possible achievement levels and reading proficiency have a greater impact on implementation in earlier or later implementation stages). While achievement levels and financial resources could fall under the umbrella of the availability of dedicated resources (which represents a predictor of implementation), it is one of many variables involved in the successful uptake of a program such as RtI.

Reading Proficiency and Intervention Intensity

Consistent with expectations, and the theoretical RtI model, intervention provision at Tier 3 (115 minutes per week per student on average) was significantly more intense than at Tier 2 (63 minutes per week per student on average). This finding is encouraging in terms of schools' capacity to provide interventions at varying intensities based on student need. In unpacking this finding, students at Tier 2 typically were in a group size of 2-3 and received 118 minutes per week of intervention. Considering these findings in the context of general recommendations provided by Brown-Chidsey & Steege (2010), Tier 2 interventions appeared to be implemented at appropriate intensity. Students in Tier 2 received approximately the recommended number of minutes per week

(120), and were actually in smaller groups, on average, than recommended (groups of 4-6) (Brown-Chidsey & Steege, 2010). At Tier 3, students were in a group size of 1-2 and received approximately 135 minutes of intervention per week. This indicated that the change in intensity from Tier 3 to Tier 2 was predominantly due to a reduction in group size. So while students at Tier 3 were receiving interventions in an appropriate group size based on recommendations (groups of 1-3 students), Tier 3 interventions may be lacking in terms of duration (number of minutes). It is recommended that students receive significantly more than 150 minutes per week, which was not the case on average (Brown-Chidsey & Steege, 2010).

Despite these expected findings, a significant relationship was not observed between achievement levels and intervention intensity at Tier 2 or Tier 3 (the significant relationship at Tier 3 was attributable to the covariate). The lack of a relationship may be due to a number of factors, however, it is hypothesized that schools may be imposing the 80-15-5 structure onto their population independent of the actual reading achievement levels. For example, although a school may have 60% of their students below proficiency and therefore considered at risk, a school could still select the lowest 15% to receive Tier 2 interventions and lowest 5% to receive Tier 3 interventions. This approach would circumvent the logistical challenges for low achieving schools, however, it would raise new questions about whether imposing the 80-15-5 model on low achieving schools can still promote successful student outcomes.

Limitations and Future Research

Extant data. The primary limitation of the study was due to the extant nature of the data/survey. The survey was conceptualized for internal purposes (at the state level)

to assess how schools were utilizing response-to-intervention to manage the specific learning disability rule that allows the use of a process of systematic evidence-based interventions. Because of this, the survey was intended to provide a snapshot of implementation and it was not originally intended for the purposes of this research study. This represents a limitation as survey development was not guided by theory (implementation science) and there also lacked some specificity in the data.

Guiding theory - implementation science. The lack of a guiding theory, in this case implementation science, represents a limitation because there are many variables thought to contribute to successful implementation of a given program such as: time, resources, leadership, strategic planning, preparation of professionals, quality of communication between the purveyor and organization, the extent to which an organization has a learning culture and central decision-making structure, and the extent to which implementation was seen as relatively easy and compatible with the organizations philosophy (Jimerson, Burns, & VanDerHeyden, 2007; Fixsen et al., 2005; Sanetti & Kratochwill, 2009). However, these variables were not assessed, and although achievement levels and school finance data were selected based on a theoretical and empirical rationale, they were also selected, in part, because of their availability.

Additionally, the use of implementation science as a guiding theory may have engendered the assessment of the implementation stages delineated by Fixsen et al. (2005). If the survey were framed in terms of implementation stages, schools in different stages could be identified, and then analyzed for differential impact of various predictors on implementation across stages. For example, in this study school finances and achievement levels were not significantly predictive of high-quality reading

interventions; however, it was hypothesized that the majority of schools were in an early stage of implementation and perhaps school finances and achievement levels significantly impact implementation at later stages of the implementation process.

Data specificity. In addition to the lack of a guiding theory underlying the survey, there were limitations in the specificity of the data available. Both independent variables, achievement level and school finance data, were limited by the data available in the state database. Specifically, the achievement data was limited by the nature of the state standardized accountability test. Although a valid proxy for school level achievement, the state standardized test does not capture achievement levels of students before grade 3, and therefore, a significant portion of the school populations was uncaptured. Further, the state standardized test was used to approximate the number of students in need of tiered services, however, the state standardized test is not meant for this purpose and school level screening data (e.g. MAP) would be better equipped for this function.

School finance data was also limited, in that it was only available at the district level. Many schools were the only elementary school within their district, which ensures that the district level data is an accurate representation of the school. Also, per-pupil expenditure varies more considerably across grade levels served (e.g. high school v. elementary), but since all schools were elementary schools, we would expect the district level data to be a good approximation for the school's funding levels. However, schools did vary to some extent in terms of the grade-levels being served, and it is possible that a level of variability of school finance data across schools within the same districts may have gone undetected.

In addition to the independent variables, there were some limitations with the survey questions that were used to create the dependent variables. Specifically, in reference to progress monitoring, respondents were not asked of the duration of progress monitoring within the context of the interventions. Based on research by Christ et al. (2013), the number of weeks that progress monitoring occurs is also a critical component contributing to the psychometric quality of the slopes of progress. Additionally, as it pertains to the questions of duration and intensity of intervention implementation, it was not explicitly stated whether the number of minutes of intervention at Tier 3 was in addition to, or in replacement of, the Tier 2 intervention minutes. Although it seems unlikely based on school capacity, it could be such that the approximately 135 minutes of Tier 3 intervention was actually in addition to the 120 minutes provided at Tier 2, which would be well beyond recommendations provided by Brown-Chidsey & Steege (2010).

Future research. The assessment of additional theoretical predictors of implementation and the assessment of implementation stages are potential avenues for future research. Data regarding school finances and achievement levels could be solicited directly from the sample schools to obtain more accurate and specific data. Additionally, the duration of progress monitoring should be collected in future research.

Self-report measures. There are limitations fundamental to studies utilizing self-report measures. Specifically, social desirability and response bias may have had an impact on the results of this study.

Social desirability. As with any study that utilizes a self-report scale, social desirability can bias responses. To the best of this author's knowledge, research on the social desirability in a similar context has not been conducted (i.e., school level personnel

reporting to their state department of education), however, there may be a tendency to over-report the ability to implement a program such as RtI. Specifically, the number of interventions being utilized, the training methodologies (often in-service and ongoing professional development), and the average intervention intensity was impressively high. However, in general, quality indicators were relatively low, and therefore, it is hypothesized that the self-report ratings were not overly positive.

Response bias. Response biases may have impacted the representativeness of the sample. Specifically, schools with limited knowledge or with limitations in their RtI implementation may have avoided completing the survey, which would create a non-representative sample. Additionally, it was noted that the sample schools generally did not represent a diverse sample of students in terms of racial/ethnic variability. Examining the participating schools indicated that there was minimal representation of schools from urban areas. This demonstrates some level of response bias that limits the inferential capacity of the sample, however, it is unclear as to why this occurred.

Future research. Future research could combine other assessment methodologies with the survey (e.g. observational assessment) to ensure that social desirability did not occur, while also ensuring the accuracy of the self-report. Additionally, schools in specific areas could be solicited for participation (i.e., particularly schools in urban areas) to ensure representativeness of the sample.

Defining, coding, and evaluating evidence-based interventions. As described above, research has established a foundation for the provision of evidence-based reading interventions, however, best research evidence may include many different research designs, including: clinical observation, qualitative designs, systematic case studies,

single-case experimental designs, ethnographic research, process-outcome studies, interventions in naturalistic settings, randomized controlled trials (RCTs), and meta-analyses, with some methods better suited to address certain types of questions (Stoiber & Desmet, 2010). The many variations of 'evidence' represents a significant challenge in both research and practice in terms of defining, selecting, coding, and evaluating evidence-based interventions, hence the many reports and review organizations that have emerged in serving this function (Stoiber & Desmet, 2010; Coffee, Newell, & Kennedy, 2014). These challenges represent a potential limitation in the study, particularly in reference to the coding of interventions. For example, in many cases, there were different ratings of evidence across the different organizations (WWC, Promising Practices, and Johns Hopkins Best Evidence Encyclopedia), interventions were often not reviewed in a given database, and there was not a pre-specified method for the consideration of newer reviews (e.g. should a newer review from one organization outweigh an older review from another organization?). All of these considerations could impact the coding scheme and therefore the quality rating of the interventions.

Not only are there challenges in the coding of evidence-based interventions, but quality may also depend on the actual implementation of the interventions. To specify, schools were provided equal value for the implementation of a given intervention and yet they may actually be implemented with different levels of quality. For example, two different schools may be using the same interventions, but the level of intervention adherence and quality of delivery may vary considerably. This alludes to a limitation of relying solely on quantitative methods, and the need to consider quality from a qualitative perspective as well.

Practical Implications and Conclusion

The findings of this study has implications for school-based program implementation in general and also as it applies to reading interventions in the context of RtI. The examination of variables known to impact the effects of implementation has received limited attention in the literature (Sanetti & Kratochwill, 2009). Accordingly, the lack of a significant relationship between school achievement levels and school finance data on the quality of reading interventions is noteworthy in-and-of-itself. This, finding, or lack thereof, is actually encouraging to some extent. The hypothesized relationship in this study was such that outdated state funding systems coupled with challenges associated with high poverty and low achievement were thought to potentially hinder the ability to implement high-quality reading interventions. However, no such relationship was found, suggesting, if nothing else, that these system-level variables (largely out of a school's control) were not strongly predictive of the ability to provide quality interventions. Conversely, other variables (under the purview of a school's control), such as school leadership, strategic planning, and a learning culture, may be the critical variables in ensuring the successful implementation of high-quality reading interventions (Jimerson, Burns, and VanDerHeyden, 2007; Fixsen et al., 2005).

That being said, the results of this study must be interpreted with caution because of the unstudied role of implementation stages and other variables that may impact implementation. As described above, it is hypothesized that many schools were in preliminary stages of implementation and the overall quality of implementation was generally low. Therefore, achievement levels and school finances may impact implementation at later stages. Additionally, it is also possible that in spite of the

variability of funding across schools, funding could be uniformly too low to enable high-quality reading intervention implementation. In other words, even among schools with greater funding, the demands of high-quality reading interventions may remain too substantial, which could also explain the generally low quality of implementation.

Collectively these results highlight the necessity for conceptualizing school-based program implementation from a theoretical perspective. This will enable an understanding of how systems-level variables differentially impact implementation across stages of implementation and ensure that schools can be appropriately supported in their implementation. In this study, a clearer understanding of implementation would enable the support of high-quality reading interventions in the context of RtI, which pertains to the larger legislative aims of ensuring equitable reading outcomes for all students.

References

Aimsweb reading. Retrieved April 10, 2014 from: http://www.aimsweb.com/products/subscriptions/reading

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

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61,* 271-285.

Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology, 51,* 1-18.

Baker, B. D., Sciarra, D. G., & Farrie, D. (2014). Is school funding fair? A national report card. Newark, NJ: Education Law Center.

Balas, E. A. & Boren, S. A. (2000). Managing clinical knowledge for health care improvement. In J. Bemmel & A. T. McCray (Eds.), *Yearbook of Medical Informatics 2000: Patient-Centered Systems* (pp. 65-70). Stuttgart, Germany: Schattauer.

Batsche, G., Elliott, J., Graden, J., Grimes, J., Kovaleski, J. F., Prasse, D., et al. (2005). *IDEA 2004 and Response to Intervention: Policy Considerations and Implementation.* Alexandria, VA: National Association of State Directors of Special Education.

Bergan, J. R. (1977). *Behavioral Consultation.* Columbus, OH: Charles E. Merrill.

Bergan, J. R. & Kratochwill, T. R. (1990). *Behavioral Consultation and Therapy.* New York: Plenum Press.

Best Evidence Encyclopedia. About the Best Evidence Encyclopedia. Retrieved April 10, 2014 from: http://www.bestevidence.org/aboutbee.htm

Brown-Chidsey, R., & Steege, M. W. (2010). *Response to intervention 2nd edition.* New York, NY: Guildford Press.

Brownson, R. C., Colditz, G. A., & Proctor, E. K. (Eds.). (2012). *Dissemination and implementation research in health.* New York, NY: Oxford University Press

Chard, D. J., Stoolmiller, M., Harn, B. A., Wanzek, J., Vaughn, S. Linan-Thompson, S., & Kame'enui, E. J. (2008). Predicting reading success in a multilevel schoolwide reading model: A retrospective analysis. *Journal of Learning Disabilities, 41,* 174-188.

Christ, T. J., Zopluoglu, C., Monaghen, B. D., & Van Norman, E. (2013). Curriculum-based measurement of oral reading: Multi-study evaluation of schedule, duration, ad dataset quality on progress monitoring outcomes. *Journal of School Psychology, 51,* 19-57.

Cochrane, W. S., & Laux, J. M. (2008). A survey investigating school psychologists' measurement of treatment integrity in school-based implementation and their beliefs about its importance. *Psychology in the Schools, 45,* 499-507.

Coffee, G., Newell, M. L., & Kennedy, A. S. (2014). Supporting collaborative efforts in implementing evidence-based reading interventions: The role of online databases. *Journal of Educational and Psychological Consultation, 24(2),* 81-95.

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology, 98,* 394-409.

Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007, January 26). The early years: Algorithm-guided individualized reading instruction. *Science, 315,* 464-465.

Crawford, L., Tindal, G., & Stieber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment, 7,* 303-323

Daly, E. J., III, Wright, J. A, Kelly, S. Q., & Martens, B. K. (1997). Measures of early academic skills: Reliability and validity with a first grade sample. *School Psychology Quarterly, 12,* 268-280.

Deno, S. (1985). Curriculum-based measurement: the emerging alternative. *Exceptional Children, 52,* 219-232.

Deno S. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37,* 184-192.

Deno, S. (2005). Problem solving assessment. In R. Brown-Chidsey (Ed.). *Assessment for intervention: A problem-solving approach* (pp. 10-42). New York: Guildford Press.

Denton, C. A., Tolar, T. D., Fletcher, J. F., Barth, A. E., Vaughn, S., & Francis, D. J. (2013). Effects of tier 3 intervention for students with persistent reading difficulties and characteristics of inadequate responders. *Journal of Educational Psychology, 105(3),* 633-648.

Durlack, J. A., & DuPre, E. P. (2008). Implementation matters: A review on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41,* 327-350.

Epstein, D. (2011). Measuring inequity in school funding. *Center for American Progress.* Retrieved April 10, 2014 from: http://www.americanprogress.org/issues/education/report/2011/08/03/10122/measuring-inequity-in-school-funding/

Equity and Excellence Commission (2013). For each and every child: A strategy for education equity and excellence. Retrieved April 10, 2014 from: http://www2.ed.gov/about/bdscomm/list/eec/equity-excellence-commission-report.pdf

Espin, C. A., Deno, S. L., Maruyama, G., & Coehn, C. (1989, March). *The Basic Academic Skills Samples (BASS): An instrument for the screening and identification of children at risk for failure in regular education classrooms.* Paper presented at the National Convention of the American Educational Research Association, San Francisco, CA.

Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature.* Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute. The National Implementation Research Network (FMHI Publication No. 231).

Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research & Practice, 16(4),* 203-212.

Forman, S. G., & Burke, C. R. (2008). Best practices in selecting and implementing evidence-based school interventions. In A. Thomas, and J. Grimes (Eds.) *Best Practices in School Psychology V* (pp. 799-812). Bethesda, MD: National Association of School Psychologists.

Forman, S. G., Serene Olin, S., Eaton Hoagwood, K., Crowe, M., & Saka, N. (2009). Evidence-based interventions in schools: Developers' views of implementation barriers and facilitators. *School Mental Health, 1,* 26-36.

Fuchs, D., Fuchs, L. S., & Stecker, P. M. (2010). The "blurring" of special education in a new continuum of general education placements and services. *Exceptional Children, 76(3),* 301-323.

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review, 21,* 45-58.

Fuchs, L. S. & Fuchs, D. (2002). What is scientifically-based research on progress monitoring. National Center on Student Progress Monitoring. Retrieved April 10, 2014 from: http://www.studentprogress.org/weblibrary.asp

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

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

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal comprehension measures. *Remedial and Special Education, 9,* 20-28.

Fuchs, D., Mock, D., Morgan, P. L., Young, P. M. (2003). Responsiveness-to-intervention: Definitions, evidence and implications for the learning disabilities construct. *Learning Disabilities Research & Practice, 18(3),* 157-171.

Gansle, K. A. & Noell, G. H. (2007). The fundamental role of intervention implementation in assessing response to intervention. In S. R. Jimerson, M. K. Burns, and A. M. VanDerHeyden (Eds.) *The Handbook of Response to Intervention: The Science and Practice of Assessment and Intervention* (pp. 244-254). New York, NY: Springer.

Garcia, S. B., Ortiz, A. A. (2008). A framework for culturally and linguistically responsive design of response-to-intervention models. *Multiple Voices for Ethnically Diverse Exceptional Learners, 11(1),* 24-41.

Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5,* 257-288.

Goss, C. L. & Brown-Chidsey, R. (2012). Tier 2 reading interventions: Comparison of reading mastery and foundations double dose. *Preventing School Failure, 56(1),* 65-74.

Green, L. W. (2008). Making research relevant: If it is an evidence-based practice, where's the practice-based evidence? *Family Practice, 25,* 20-24.

Greenhalgh, T., Robert, G., MacFarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: Systematic review and recommendations. *The Milbank Quarterly, 82,* 581-629.

Gresham, F. M. (2014). Measuring and analyzing treatment integrity data in research. In L. H. Hagermoser Sanetti & T. R. Kratochwill (Eds.) *Treatment Integrity: A foundation for evidence-based practice in applied psychology* (pp.109-130) . American Psychological Association: Washington, DC

Hagermoser Sanetti, L. H., & Kratochwill, T. R. (2005). Treatment integrity assessment within a problem-solving model. In R. Brown-Chidsey (Ed.) *Assessment for Intervention: A Problem-Solving Approach* (pp. 304-325). New York, NY: Guilford Press.

Hagermoser Sanetti, L. M. & Kratochwill, T. R. (2008). Treatment integrity in behavioral consultation: Measurement, promotion, and outcomes. *International Journal of Behavioral Consultation and Therapy, 4,* 95-114.

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

Hagermoser Sanetti, L. M. & Kratochwill, T. R. (Eds.). (2014). *Treatment integrity: A foundation for evidence-based practice in applied psychology.* American Psychological Association: Washington, DC

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* New York, NY: Routledge.

Hintze, J. M, & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. *School Psychology Review, 34,* 372-386.

IDEIA (2004). Individuals with Disabilities Education Improvement Act. Pub. L 108-446.

ICEBeRG (2006). Designing theoretically informed implementation interventions. *Implementation Science, I,* 4.

Institute for Education Sciences (2009). *Assisting students struggling with reading: Response to intervention and multi-tier intervention in the primary grades.* Washington, DC: Author.

Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (2007). Response to intervention at school: The science and practice of assessment and intervention. In S. R. Jimerson, M. K. Burns, and A. M. VanDerHeyden (Eds.) *The Handbook of Response to Intervention: The Science and Practice of Assessment and Intervention* (pp. 3-9). New York, NY: Springer.

Joyce, B., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Kranzler, J. H., Brownell, M. T., & Miller, M. D. (1998). The construct validity of curriculum-based measurement of reading: An empirical test of a plausible rival hypothesis. *Journal of School Psychology, 36,* 399-415.

Kratochwill, T. R., (2007). Preparing psychologists for evidence-based school practice: Lessons learned and challenges ahead. *American Psychologist, 62(8),* 829-843.

Kratochwill, T. R., Clements, M. A., & Kalymon, K. M. (2007). Response to intervention: Conceptual and methodological issues in implementation. In S. R. Jimerson, M. K. Burns, and A. M. VanDerHeyden (Eds.) *The Handbook of*

Response to Intervention: The Science and Practice of Assessment and Intervention (pp. 25-52). New York, NY: Springer.

Kratochwill, T. R., & Stoiber, K. C. (2002). Evidence-based interventions in school psychology: Conceptual foundations of the procedural and coding manual of Division 16 and the Society for the Study of School Psychology Task Force. *School Psychology Quarterly, 17,* 341-389.

Kovaleski, J. F. (2007). Potential pitfalls of response to intervention. In S. R. Jimerson, M. K. Burns, and A. M. VanDerHeyden (Eds.) *The Handbook of Response to Intervention: The Science and Practice of Assessment and Intervention* (pp. 80-92). New York, NY: Springer.

Marston, D. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). New York: Guildford Press.

McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review, 33,* 193-203.

National Center for Research Statistics. (2013). *A first look: 2013 mathematics and reading.*

National Center on Response to Intervention (2010). *Essential components of RtI – A closer look at response to* intervention. Retrieved April 10, 2014 from http://rti4success.org

Noell, G. H. (2010). Empirical and pragmatic issues in assessing and supporting intervention implementation in schools. In S. R. Jimerson, M. K. Burns, and A.

M. VanDerHeyden (Eds.) *The Handbook of Response to Intervention: The Science and Practice of Assessment and Intervention* (pp. 25-52). New York, NY: Springer.

Noell, G. H. & Gansle, K. A. (2014). The use of performance feedback to improve intervention implementation in schools. In L. H. Hagermoser Sanetti & T. R. Kratochwill (Eds.) *Treatment Integrity: A foundation for evidence-based practice in applied psychology* (pp.161-183) . American Psychological Association: Washington, DC

Noell, G. H., & Gansle, K. A. (2006). Assuring the form has substance: Treatment plan implementation as the foundation of assessing response to intervention. *Assessment for Effective Intervention, 32,* 32-39.

Noell, G. H., Witt, J. C., Gilbertson, D. N., Ranier, D. D., & Freeland, J. T. (1997). Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *School Psychology Quarterly, 12,* 77-88.

Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., & Williams, K. L.… Duhon, G. J. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review, 34,* 87-106.

Ogden, T. & Fixsen, D. L. (2014). Implementation science: A brief overview and a look ahead. *Zeitschrift fur Psycholgie, 222(1),* 4-11.

Perry, C. L., Murray, D. M., & Griffin, G. (1990). Evaluating the stateside dissemination of smoking prevention curriculua: Factors in teacher compliance. *Journal of School Health, 60,* 501-504.

Promising Practices Network. How programs are considered. Retrieved April 10, 2014 from http://www.promisingpractices.net/criteria.asp

Ransford, C. R., Greenberg, M. T., Domitrovich, C. E., Small, M., & Jacobson, L. (2009). The role of teachers' psychological experiences and perceptions of curriculum supports on the implementation of a social and emotional learning curriculum. *School Psychology Review, 38(4),* 510-532.

Reading recovery council of North America. Retrieved April 10, 2014 from: http://readingrecovery.org/reading-recovery/training/for-teachers

Requirements and costs for the training of reading recovery in Ohio, 2013. Retrieved April 10, 2014 from: http://www.rrosu.org/Becoming/REQUIREMENTS_AND_COSTS_Amended_Apr_2013.pdf

Reschly, D. J., (2008). School psychology paradigm shift and beyond. In A. Thomas, and J. Grimes (Eds.) *Best Practices in School Psychology V* (pp. 3-15). Bethesda, MD: National Association of School Psychologists.

Ross, J. G., Leupker, R. V., Nelson, G. D., Saavedra, P., & Hubbard, B. M. (1991). Teenage health teaching modules: Impact of teacher training on implementation and student outcomes. *Journal of School Health, 61,* 31-38.

Shakow, D., Hilgard, E. R., Kelly, E. L., Luckey, B., Sanford, R. N., & Shaffer, L. F. (1947). Recommended graduate training program in clinical psychology. *American Psychologist, 2,* 539-558.

Shapiro, E. S., (2013). Commentary on progress monitoring with CBM-R and decision making: Problems found and looking for solutions. *Journal of School Psychology, 51,* 59-66.

Shapiro, E. S., Zogmond, N., Wallace, T., & Marston, D. (2011). *Models for Implementing Response to Intervention: Tools, Outcomes, and Implications*. New York, NY: Guilford Press.

Shinn, M. R. (2008). Best practices in using curriculum-based measurement in a problem-solving model. In A. Thomas, and J. Grimes (Eds.) *Best Practices in School Psychology V* (pp. 243-262). Bethesda, MD: National Association of School Psychologists.

Silberglitt, B., & Hintze, J. (2005). Formative assessment using CBM-R cut scores to track progress toward success on state mandated achievement tests: A comparison of methods. *Journal of Psychoeducational Assessment, 23,* 304-325.

Simmons, D. C., Kame'enui, E. J., Coyne, M. D., & Chard, D. J. (2002). Effective strategies for teaching beginning reading. In E. Kame'enui, D. Carmine, R. C. Dixon, D. C. Simmons, & Coyne (Eds.), Effective teaching strategies that accommodate diverse learners (2nd ed., pp 53-992). Columbus, OH: Merrill/Prentice Hall.

Slavin, R. (2008). What works? Issues in synthesizing educational program evaluations. *Educational Researcher, 37 (1),* pp. 5-14.

Stecker, P., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42,* 795-819.

Stoiber, K. C., DeSmet, J. L. (2010). Guidelines for evidence-based practice in selecting interventions. In G. G. Peacock, R. A. Ervin, E. J. Daly III, & K. W. Merrell (Eds.) *Practical Handbook of School Psychology: Effective Practices for the 21st Century* (pp. 213-234). New York, NY: Guilford Press.

Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics (5th ed.).* Boston, MA: Pearson.

Upah, K. R. F. (2008). Best practices in designing, implementing, and evaluating quality interventions. In A. Thomas, and J. Grimes (Eds.) *Best Practices in School Psychology V* (pp. 209-224). Bethesda, MD: National Association of School Psychologists.

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

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45,* 225-256.

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,* 126-142.

Wayman, M. M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education, 41(2),* 85-120.

What Works Clearinghouse, 2010. Procedures and Standards Handbook Version 2.1. Retrieved April 10, 2014 from http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=19

Witt, J. C., Noell, G. H., LaFleur, L. H., & Mortenson, B. P. (1997). Teacher usage of interventions in general education: measurement and analysis of the independent variable. *Journal of Applied Behavior Analysis, 30,* 693-696.

U.S. Department of Education (2001). The Elementary and Secondary Education Act (The No Child Left Behind Act of 2001).

Zirkel, P. A. (2012) *The legal dimension of RTI: Part II. State laws and guidelines.* Retrieved April 10, 2014 from http://rtinetwork.org

Appendix A

Survey Questions and Response Options

Directions: When responding to the following questions, please consider only those practices in which your school currently engages. Please select those areas of instruction for which you currently utilize the interventions or progress monitoring tools listed. Though you may select any of the eight areas listed for each intervention or progress monitoring tool listed, it is not expected that all will be applicable. For example, it is expected that math interventions will only be used for mathematics calculation and/or mathematics problem solving, etc.

Evidence-Based Interventions

1. *At this point in time*, what interventions are used at your school for students receiving more intensive instruction/*Tier 2* instruction, for each of the following reading/language arts?

Accelerated Reader

Basal Reader Series (Title & Publisher)
Compass Learning Odyssey
Corrective Reading (SRA)
Drill/Flashcards
Early Intervention in Reading (EIR)
Guided Reading
Individualized Instruction/Tutoring
Leveled Literacy Intervention (Fountas & Pinnell)
Orton-Gillingham
Partner Reading
Peer Tutoring
Read 180
Read Naturally
Readers' Theatre
Reading Mastery
Reading Recovery
Reciprocal Teaching
Self-Regulated Strategy Instruction
SOAR to Success (Houghton-Mifflin
Wilson Reading System
Other (Write-in)

2. *At this point in time*, what interventions are used at your school for students receiving more *Tier 3* instruction, for each of the following reading/language arts?

Accelerated Reader

Basal Reader Series (Title & Publisher)
Cognitive Strategy Instruction
Compass Learning Odyssey
Corrective Reading (SRA)
Drill/Flashcards
Early Intervention in Reading (EIR)
Guided Reading
Individualized Instruction/Tutoring
Leveled Literacy Intervention (Fountas & Pinnell)
Orton-Gillingham
Partner Reading
Peer Tutoring
Read 180
Read Naturally
Readers' Theatre
Reading Mastery
Reading Recovery
Reciprocal Teaching
Self-Regulated Strategy Instruction
SOAR to Success (Houghton-Mifflin
Wilson Reading System
Other (Write-in)

Progress Monitoring

3. *At this point in time*, what progress monitoring tools does your school use for students receiving tier 2 instruction, for reading/language arts?

AIMSWeb

Basal Reading Series

Curriculum Based Measures –
Reading (CBM-R)
Dynamic Indicators of Basic
Early Literacy Skills (DIBELS)
Developmental Reading
Assessment (DRA)
Easy-CBM
Reading Recovery Assessments
Running Records
STAR Reading
Other (Write-in)

4. *At this point in time*, what progress monitoring tools does your school use for students receiving tier 3 instruction, for reading/language arts?

AIMSWeb

Basal Reading Series
Curriculum Based Measures –
Reading (CBM-R)
Dynamic Indicators of Basic
Early Literacy Skills (DIBELS)
Developmental Reading
Assessment (DRA)
Easy-CBM
Reading Recovery Assessments
Running Records
STAR Reading
Other (Write-in)

Treatment Integrity

5. *At this point in time*, how is your school monitoring implementation fidelity for the tier 2 intervention being provided to students in reading/language arts?

Checklist

Video-taping
Peer feedback
Outside monitoring
Other (Write-in)

6. *At this point in time*, how is your school monitoring implementation fidelity for the tier 3 intervention being provided to students in reading/language arts?

Checklist

Video-taping
Peer feedback
Outside monitoring
Other (Write-in)

Trained Interventionist

7. *At this point in time,* who administers these tier 2 interventions in reading/language arts?

Classroom teacher(s)

Curriculum specialist
Interventionist
Literacy coach
Reading specialist
Reading teacher
Para-professional
School psychologist
School counselor
Special education assistant
Special education teacher
Title 1 teachers
Other (write-in)

8. Who was included in these training sessions for reading/language arts?

Classroom teacher(s)

Curriculum specialist
Interventionist
Literacy coach
Reading specialist
Reading teacher
Para-professional
School psychologist
School counselor
Special education assistant
Special education teacher
Title 1 teachers
Other (write-in)

9. *At this point in time,* how has your school developed the capacity to implement the interventions that are used with students receiving tier 2 instruction in reading/language arts?

On-going professional development
In-service day(s)

Materials introduced at a faculty meeting
Materials introduced at a team meeting
No training has been provided to date

10. *At this point in time,* who administers these tier 3 interventions in reading/language arts?

Classroom teacher(s)

Curriculum specialist
Interventionist
Literacy coach
Reading specialist
Reading teacher
Para-professional
School psychologist
School counselor
Special education assistant
Special education teacher
Title 1 teachers
Other (write-in)

11. Who was included in these training sessions for reading/language arts?

Classroom teacher(s)

Curriculum specialist
Interventionist
Literacy coach
Reading specialist
Reading teacher
Para-professional
School psychologist
School counselor
Special education assistant
Special education teacher
Title 1 teachers
Other (write-in)

12. *At this point in time,* how has your school developed the capacity to implement the interventions that are used with students receiving tier 3 instruction in reading/language arts?

On-going professional development

In-service day(s)
Materials introduced at a faculty meeting
Materials introduced at a team meeting
No training has been provided to date

Appropriate Group Size

13. How are these tier 2 interventions typically provided to students for reading/language arts?

Individualized instruction

Small group of 2-3 students
Small group of 3-5 students
Small group of 6 or more students

14. How are these tier 3 interventions typically provided to students for reading/language arts?

Individualized instruction

Small group of 2-3 students
Small group of 3-5 students
Small group of 6 or more students

Appropriate Intervention Duration

15. How often are these tier 2 interventions typically provided to students for reading/language arts?

	How many sessions per week						How long per session			
Reading/language arts	1	2	3	4	5	5+	< 15	15-30	31-45	> 45

16. How often are these tier 3 interventions typically provided to students for reading/language arts?

	How many sessions per week						How long per session			
Reading/language arts	1	2	3	4	5	5+	< 15	15-30	31-45	> 45

Appropriate Frequency of Progress Monitoring

17. How often is progress monitoring conducted for students receiving tier 2 instruction in reading/language arts?

Daily

4 times/week
3 times/week
2 times/week
Weekly
Every 2 weeks
Once/month

18. How often is progress monitoring conducted for students receiving tier 3 instruction in reading/language arts?

Daily

4 times/week
3 times/week
2 times/week
Weekly
Every 2 weeks
Once/month

Appendix B

Intervention Frequency by Tier

	Tier 2	Tier 3	Total
Accelerated Reader	12	2	14
Compass Learning Odyssey	7	9	16
Corrective Reading	9	20	29
Early Intervention in Reading	8	32	40
Leveled Literacy Intervention (Fountas & Pinnell)	56	10	66
Orton-Gillingham	8	6	14
Read 180	8	8	16
Read Naturally	15	4	19
Readers' Theatre	13	4	17
Reading Mastery	5	10	15
Reading Recovery	11	2	13
Reciprocal Teaching	3	0	3
SOAR to Success (Houghton-Miffin)	12	2	14
Wilson Reading System	4	8	12
Other			
Voyager	2	2	4
Imagine Learning	1	1	2
Lexia	1	1	2
PALS	1	1	2
Sound Partners	1	1	2
Early reading empowerment	2	2	4
General practices			
Drill/Flashcards	30	6	36
Guided Reading	59	38	97
Individualized Instruction/Tutoring	36	40	76
Partner Reading	23	8	31
Peer Tutoring	6	7	13

CURRICULUM VITAE

EDUCATION

PhD in School Psychology (Summer 2015; anticipated)
University of Wisconsin-Milwaukee (APA Accredited)

MS in Educational Psychology (2011)
University of Wisconsin-Milwaukee

BS in Psychology (2009)
University of Wisconsin-Madison

SCHOOL-BASED EXPERIENCE

Lawrence D. Crocker College Prep (07/2014 – present)
Supervisor: Dr. Kelli Jordan
Responsibilities:

Response-to-Intervention Coordinator

- Developed formal academic screening process using multiple measures
- Established measures and systems for monitoring the integrity of intervention implementation
- Created a developmental progress monitoring sequence for students receiving academic interventions
- Coordinated the implementation of multiple skills/behavior groups, check-in check-out, and behavior plans
- Implemented multiple academic intervention groups and skills/behavior groups (e.g. anger group, social skills group)
- Provided behavioral consultation for teachers and support staff

Evaluation Coordinator

- Completed Bulletin 1508 compliant evaluations (to date: 22)
- Responsible for all components of the evaluation
    - Interviews, observations, psychological assessment, educational assessment
- Responsible for the integrated report, staffing/eligibility determination, and dissemination

Milwaukee Spanish Immersion School & Vincent High School (09/2010 – 06/2011)
Supervisor: Travis Pinter, Ed.S., and Dr. Edgar Jordan
Responsibilities:

- Progress monitoring for over 25 students at-risk for reading difficulties
- Conducted psychoeducational evaluations for Special Education, 504, and Non-Special Education Need referrals
- Facilitated academic and mental health interventions
- Provided individual therapy utilizing cognitive behavioral methods and participated in Restorative Justice to help students avoid suspensions
- Individualized consultation utilizing a problem–solving approach

- Experience with crisis response teams (short-term individual and group grief counseling)

Milwaukee Public Schools Dept. of Research & Evaluation (09/2011 – 06/2012)
Supervisor: Marc Sanders
Responsibilities:

- Engaged in district-wide research, which includes such responsibilities as:
  - Consultation with various outside vendors evaluating district initiatives or programs
  - Evaluating corrective action requirements
  - Grant writing development (e.g. Race to the Top, Gear-Up, and Gear-Up data utilization grant)
- Provided technical assistance and presented data at the school and district level to relevant stakeholders
- Developed methods to ensure the fidelity of assessment and curriculum implementation

CLINICAL EXPERIENCE

Family Options Counseling (12/2012 – 12/2013)
Supervisor: Dr. Kimberly Young
Responsibilities:

- Provided therapeutic services at the individual, group and family level often with a trauma focus
  - Provided individual therapy utilizing behavioral and cognitive-behavioral techniques within an ecological perspective
  - Co-facilitated 3 groups focusing on reducing the risk of sexual offense recidivism among children and adolescents with cognitive deficits
  - Co-facilitated an adolescent anger management group using the aggression replacement training approach for adolescents
  - Co-facilitated a life skills group for students with cognitive deficits who have graduated high school
- Conducted psychological evaluations

Medical Colleges of Wisconsin – Neuropsychology (04/2013 – 06/2014)
Supervisor: Dr. Jennifer Koop / Dr. Amy Heffelfinger
Responsibilities:

- Conducted neuropsychological assessments
- Managed two research databases: the Pediatric Epilepsy and Pediatric Brain Tumor databases.
  - Coding, entry, and organization of these datasets
  - Preliminary analyses involving neuropsychological outcomes for pediatric epileptic patients who have undergone a hemispherectomy and children with a posterior fossa tumor who have undergone a resection

RESEARCH EXPERIENCE

Research Assistant / Project Manager (09/2010 – 06/2014)
University of Wisconsin-Milwaukee

Supervisor: Dr. Markeda Newell
Responsibilities:

- Assisted in the development of a computer simulation used to study consultation training.
- Conducted multiple research projects and wrote eventual manuscripts
- Consulted with the Wisconsin Response to Intervention (RtI) center in the development of Culturally-Responsive Practices
- Collaborated with the School Psychology Consultant of the Wisconsin Department of Public Instruction, to develop a Pupil Services Multicultural Survey for the state of Wisconsin

Research Assistant (01/2009 – 05/2009)
University of Wisconsin-Madison
Department of Social Psychology
Supervisor: Dr. Judith Harackiewicz
Responsibilities:

- Conducted research on motivation, goal-setting and competition paradigms and the effects on performance
- Involved in participant recruitment, data analysis, and implementation of the research paradigm

TEACHING EXPERIENCE

Instructor (Spring 2014)
Educational Psychology: The Practice of Classroom Assessment
Responsibilities:

- Developed curriculum and lectures on theoretical frameworks, assessment development (i.e. creation of assessments), and trends in assessment as applied to teachers

Teaching Assistant (Fall 2012 – Spring 2014)
Assessment and Intervention: School Age Children, Academic Interventions and Alternative Assessment
Supervisor: Dr. Kyongboon Kwon
Responsibilities:

- Collaborated on curriculum and activity development
- Lectured on various topics related to standardized, individually administered cognitive tests.
- Evaluated student work, including video-taped test administrations, worksheets, exams and protocols

PUBLISHED MANUSCRIPTS

Looser, J. A. (2013). Guidelines for the prevention, detection, and follow-up of high-stakes testing irregularities. *National Association for School Psychologist's, Communiqué.*

Newell, M.L, Newell, T.S., & **Looser, J. A.** (2013). A competency-based assessment of school-based consultants' implementation of consultation. *Training and Education in Professional Psychology*.

Newell, M.L., Newell, T.S., & **Looser, J. A.** (2013). Examining how novice consultants address cultural factors during consultation: Illustration of a computer-simulated case-study method. *Consulting Psychology Journal: Practice and Research, Vol 65(1)*, 74-86.

Newell, M.L. & **Looser, J. A.** (2013). Educational Achievement (of different ethnic groups). In Multicultural America: A Multimedia Encyclopedia. Thousand Oaks, CA: Sage Publications.

PRESENTATIONS

Looser, J. A., & Newell, M. L. (2014, February). *Multicultural competence: A survey of pupil service providers.* Presented at the annual convention of the National School Psychologist Association, Washington D.C.

Callan, G.L., Cleary, T.J., Reynolds, C.E., **Looser, J.A.**, Schumaker, C., Rollo, K. (2014, February). *Self-regulated learning microanalysis for math problem-solving.* Presented at the annual convention of the National School Psychologist Association, Washington D.C.

Callan, G., Cleary, T., Reynolds, C., & **Looser, J. A.** (2013, May). *Measuring self-regulated learning (SRL) during mathematical problem solving with SRL microanalysis.* Presented at the annual University of Wisconsin-Milwaukee Doctoral Student Research Session, Milwaukee, WI.

Sanders, M., **Looser, J. A**., (2013, October). *Practical application of testing irregularities in a large urban district.* Presented at the Statistical Detection of Potential Test Fraud Conference, Madison, WI.

AWARDS/AFFILIATIONS

Chancellor's Graduate Student Award (2012 & 2013)

University of Wisconsin-Madison, Dean's List (2006-2009)

Graduate Student Representative of Wisconsin School Psychology Association

National Association of School Psychologists (Student Member)

UWM School Psychology Student Association (Student Member)