2016

Evaluating the Implementation Quality and Utility of Response to Intervention Practices

Paige Hamilton-Read
University of Rhode Island, phamilton318@gmail.com

Follow this and additional works at: https://digitalcommons.uri.edu/oa_diss

Recommended Citation

Hamilton-Read, Paige, "Evaluating the Implementation Quality and Utility of Response to Intervention Practices" (2016). *Open Access Dissertations*. Paper 522.
https://digitalcommons.uri.edu/oa_diss/522

This Dissertation is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu.

EVALUATING THE IMPLEMENTATION QUALITY AND UTILITY OF
RESPONSE TO INTERVENTION PRACTICES

BY

PAIGE HAMILTON-READ

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
PSYCHOLOGY

UNIVERSITY OF RHODE ISLAND

2016

DOCTOR OF PHILOSOPHY

OF

PAIGE HAMILTON-READ

APPROVED

Dissertation Committee:

Major Professor

Gary Stoner

Susan Loftus-Rattan

Julie Coiro

Nasser H. Zawia

Dean Of The Graduate School

UNIVERSITY OF RHODE ISLAND

2016

ABSTRACT

The current study investigated the extent to which eight elementary schools in one cooperative education district implemented the five essential components of Response to Intervention (RtI), and explored the relationship between fidelity of implementation and student reading outcomes measured by oral reading fluency (ORF). Various RtI models exist because there is no single method to implement RtI appropriately. The majority of available studies examining fidelity of RtI implementation have focused on the individual components of RtI. However, when implemented as intended, RtI is a coherent system of coordinated components. Consequently, it is important to study the implementation of RtI as a whole model in addition to the individual components and its' relationship with student outcomes. Descriptive statistics were used to analyze the extent of implementation and to characterize differences between the schools and analysis of variance (ANOVA) was used to examine differences in student outcomes between elementary schools. Finally, the researcher made qualitative inferences to explore the relationship between fidelity of implementation and student reading outcomes measured by ORF. Results from the current study revealed that fidelity of RtI implementation varied between elementary schools despite similar professional development and supports around RtI practices. In addition, results from the current study have preliminary implications supporting the value of *infrastructure and supports* and *fidelity and evaluation* within the RtI model.

ACKNOWLEDGEMENTS

This dissertation could not have been completed without the contribution of many individuals. First, my greatest gratitude goes to my major professor Dr. Gary Stoner. His guidance, mentoring, and support were invaluable during my graduate school experience, and especially this dissertation. I am grateful and fortunate that he agreed to be my advisor. Second, I want to thank my committee members, Drs. Susan Loftus-Rattan and Julie Coiro for their thoughtful feedback and support. Third, I want to thank the school professionals who provided me with the extant data set and for their continued communication with me while I wrote this dissertation. Fourth, thank you to my parents and my in-laws for their optimism and endless belief in me. Finally, to my husband, Alex, thank you for your patience and for being my side through all of the highs and lows. You more than anyone know what a crazy roller coaster ride this experience has been and I'm lucky to have you.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iii
TABLE OF CONTENTS.....	iv
LIST OF TABLES.....	vi
LIST OF FIGURES.....	viii
STATEMENT OF THE PROBLEM.....	1
JUSTIFICATION AND SIGNIFICANCE OF THE PROBLEM.....	2
Implementation Science.....	2
Response to Intervention/Instruction.....	5
Response to Intervention/Instruction Essential Components	10
Relationship Between Implementation and Outcomes.....	12
Purpose of the Study	14
METHODOLOGY.....	17
Procedures.	17
Measures.....	19
Response to Intervention/Instruction.....	19
Student Outcomes.....	26
Oral Reading Fluency.....	26
Participants.....	27
School Characteristics	27
Interviewees.....	31

TABLE OF CONTENTS (continued)

Students.....	31
Informed Consent.....	32
Design	33
RESULTS.....	34
Analyses.....	34
First Research Question.....	35
Second Research Question.....	42
DISCUSSION.....	60
How Are These Results Similar to and Different From Those of Previous Studies?.....	62
Limitations.....	68
Implications for Practice	71
Future Directions.....	81
APPENDIX A: Edited Research Question.....	84
APPENDIX B: RtI Essential Components Worksheet	86
APPENDIX C: RtI Fidelity of Implementation Rubric.....	103
APPENDIX D: RtI Measure Reliability.....	112
APPENDIX E: Normality Information and Comparison of Parametric and Nonparametric Test Results ...	114
BIBLIOGRAPHY.....	119

LIST OF TABLES

TABLE	PAGE
Table 1. Demographic Characteristics by School	29
Table 2. Total Number of Participants by School and Grade	32
Table 3. Descriptive Statistics for the RtI Measure.....	36
Table 4. Ratings on the RtI Measure by School	37
Table 5. Mean and Standard Deviations on the Measure of Spring WRC by School and Grade Level	39
Table 6. Mean and Standard Deviations on the Measure of ROI by School and Grade Level	41
Table 7. Grade 1 WRC: Results from the Kruskal-Wallis Test	45
Table 8: Grade 1 WRC: Mean Ranks for Each School Based on Mean Scores.....	45
Table 9: Grade 1 ROI: One-Way Analysis of Variance of Students' ROI by School..	46
Table 10: Grade 1 ROI: Summary of Significant Multiple Comparisons between Schools (Tukey HSD).....	46
Table 11: Grade 2 WRC: One-Way Analysis of Variance of Students' WRC by School.....	47
Table 12: Grade 2 WRC: Summary of Significant Multiple Comparisons between Schools (Tukey HSD).....	47
Table 13: Grade 2 ROI: One-Way Analysis of Variance of Students' ROI by School.....	48
Table 14: Grade 2 ROI: Summary of Significant Multiple Comparisons between Schools (Tukey HSD).....	49
Table 15: Grade 3 WRC: One-Way Analysis of Variance of Students' WRC by School.....	49
Table 16: Grade 3 WRC: Summary of Significant Multiple Comparisons between Schools (Tukey HSD).....	50

TABLE	PAGE
Table 17: Grade 3 ROI: One-Way Analysis of Variance of Students' ROI by School.....	50
Table 18: Grade 3 ROI: Summary of Significant Multiple Comparisons between Schools (Tukey HSD).....	51
Table 19: Grade 4 WRC: One-Way Analysis of Variance of Students' ROI by School.....	52
Table 20: Grade 4 WRC: Summary of Significant Multiple Comparisons between Schools (Tukey HSD)	52
Table 21. Grade 4 ROI: Results from the Kruskal-Wallis Test.....	53
Table 22. Grade 4 ROI: Mean Ranks for Each School Based on Mean Scores.....	53
Table 23. Grade 5 WRC: One-Way Analysis of Variance of Students' WRC by School.....	54
Table 24. Grade 5 WRC: Summary of Significant Multiple Comparisons between Schools (Tukey HSD)	54
Table 25. Grade 5 ROI: One-Way Analysis of Variance of Students' ROI by School.....	55
Table 26. Grade 5 ROI: Summary of Significant Multiple Comparisons between Schools (Tukey HSD)	55
Table 27. Comparison Of Significant Post-hoc Tukey Tests and Fidelity of RtI Implementation.....	59

LIST OF FIGURES

FIGURE	PAGE
Figure 1. Visual Representation of the Structure of the Education District.....	28
Figure 2. Mean Assessment Ratings by School.....	38
Figure 3. Mean DBDM Ratings by School.....	38
Figure 4. Mean MTSS Ratings by School.....	39
Figure 5. Mean Infrastructure and Support Ratings by School.....	40
Figure 6. Mean Fidelity and Evaluation Ratings by School	41

Evaluating the implementation quality and utility of Response to Intervention practices

Statement of the Problem

Response to Intervention/Instruction (RtI) is a complex and comprehensive framework for providing multiple levels of service delivery to meet individual student needs. Since 2004, an increase in the adoption of RtI has been noted in schools across the country (Berkeley, Bender, Peaster, Saunders, 2009). The increase can, in part, be attributed to federal regulations. The Individuals with Disabilities Education Improvement Act of 2004 (IDEIA, 2004) incorporated a new regulation that permits local education agencies to consider a student's response to scientifically based intervention as a procedure for determining eligibility for special education services (P.L. No. 108-446 614 [b][6][A]; 614 [b][2&3]). This provision is well aligned with the former No Child Left Behind Act (2001), which required the use of standardized assessments to inform decisions about student learning at the school wide level (No Child Left Behind [NCLB], 2002). As a result, RtI is associated with high-stakes decisions, accountability for the outcomes of instruction, and data based decision making.

Various RtI models exist because there is no single method to implement RtI appropriately. Rather, a growing body of literature has suggested that a number of central components are critical to defining an RtI model (Noltemeyer, Boone, & Sansosti, 2014; Stoiber, 2014). The National Center for Response to Intervention (2014) identified the five essential components of RtI including a) assessments, b) data-based decision making, c) multilevel instruction, d) infrastructure and support

mechanisms, and e) fidelity of implementation and evaluation. The interrelatedness of these components forms the larger RtI framework for making valid decisions about student learning (Kovaleski, 2007). In other words, a complete (and likely to be effective) RtI framework can be considered to be in place only when evidence exists to support the implementation of each of the five components.

The current study examined the extent to which RtI as a whole, and each essential component, was implemented in eight elementary schools and, additionally, explored the relationship between the quality of implementation of RtI and school level outcomes.

The current study examined the following primary research questions:

1. To what extent are staff in elementary schools implementing the essential components of RtI with integrity, as measured by the RtI Essential Components Worksheet between schools; and if differences exist between schools, how can those differences be characterized?
2. To what extent does the integrity of overall RtI implementation, and the implementation of the five essential components, in elementary schools, relate to differences student-level student outcomes, at each grade across schools, as measured by AIMSWEB screening data?

Justification and Significance of the Problem

Implementation Science

One of the most critical issues in educational research is a gap between what is known about effective practices (i.e. evidence-based practices) and how the practices are implemented in schools. While there is a growing supply of evidence-based practices in education, there is much less research evidence regarding how to support the successful implementation of these practices in school settings (Durlak & DuPre, 2008). With the increasing demand by federal regulations for the use of evidence-

based practices, it is important to understand and reduce the gap between what is known and what is occurring. When a new practice or program is studied in the natural, desired setting, its implementation can be planned and studied using methods from the area of research/scholarship known as ‘implementation science’. Resulting information can yield important conclusions regarding the utility of the practice within the context and culture of individual schools and levels of adaptation that schools may need to maintain the new practice (Shulte, Easton, & Parker, 2009).

Implementation science research has demonstrated that when evidence-based practices are implemented in schools, varying levels of implementation tend to emerge (Odom, 2009). For example, Odom et al. (2010) conducted a national study examining the implementation of a curriculum across multiple sites at seven times during the school year and found significant differences in the quality of implementation between sites. More specifically, dramatically low implementation levels were found in schools with high percentages of minority students. This study demonstrates that school characteristics may influence how new practices are implemented in schools.

According to the American Psychological Association (2005), adaptation is a natural part of implementing evidence-based practices when an individual with expertise in the area manages the modifications. Further, recent meta-analyses indicate that full implementation may not be necessary to obtain positive outcomes, but instead core components may be identified as the pathway to optimal implementation (Durlak & DuPre, 2008). Individuals who implement evidence-based practices should be knowledgeable about how to implement the practice and how to implement it with high quality. Notably, Greenwood (2009) proposes that interventions are the result of

the highly skilled behaviors of the individuals who implement them as they follow a set of manualized procedures. Examining the fidelity of program implementation is one method for evaluation in implementation science research.

Treatment Integrity. When implementing a new program it is important to understand both the extent to which it is being implemented as intended, and to understand implementation in a manner that links the program to student outcomes. That is to say, outcomes cannot confidently be attributed to a new program unless data are collected documenting how the program was implemented (Sanetti & Kratochwill, 2010). The terms treatment integrity, treatment fidelity, and fidelity of implementation are often used interchangeably in the field of education. For the purposes of the current study, the term fidelity of implementation will be used and defined as, “the extent to which essential intervention or program components are delivered in a comprehensive and consistent manner by an interventionist trained to deliver the intervention” (Hagermoser, Sanetti & Kratochwill, 2009 p. 448). Adherence and quality of implementation are the two dimensions of fidelity of implementation commonly used when evaluating a program, or framework, such as RtI. Adherence refers to identifying whether something occurred or did not occur and quality refers to the skill with which it was performed (Noell & Gansle, 2006; Schulte, Easton, & Parker, 2009). For example, if all students were assessed using a screening assessment then, on an adherence measure the school would be considered to have implemented screening; however, on a quality measure the extent to which standardized procedures were followed, and the number of times per year students are assessed would determine *how well* the school is implementing screening. Evaluating the quality of

implementation permits a more detailed analysis aimed at understanding relationships between different aspects/levels of implementation with student outcomes and it broadens the understanding of what elements are necessary to achieve improved student outcomes. Additionally, understanding the context (i.e. child engagement, classroom context, teacher involvement) in which implementation occurs is believed to improve the validity of decisions regarding program effectiveness (Greenwood & Kim, 2012).

Response to Intervention/Instruction

The primary issues noted regarding implementation science and the importance of fidelity of implementation for valid data-based decision making are relevant to the contemporary area of educational practices known as Response-to-Intervention (RtI). Since IDEIA (2004) language permitted the use of RtI to be considered when making decisions about special education eligibility, RtI has been adopted at an increasing rate in schools across the country. RtI is a comprehensive multi-tiered framework intended to improve outcomes for all students through prevention and early intervention efforts. RtI is composed of three, or more, tiers of service delivery, which are defined by the intensity, duration, and degree of individualization of service provided (i.e. method of service delivery)(Jimerson, Burns, & VanDerHeyden, 2015). Educators collaborate and engage in data based decision making to appropriately allocate instructional resources and other support resources to ensure individual student needs are met (Burns & Gibbons, 2008). Brown-Chidsey (2007) argues that RtI meets student needs in a more time efficient manner than prior models and fewer children will require special education services when RtI is implemented with fidelity. In fact, one study

found a reduction from 4.5% to 2.5% in special education placements rates over a 10-year period of RtI implementation (Bollman, Silberglitt, & Gibbons, 2007). For a detailed description of RtI, the reader is referred to Burns and Gibbons (2008).

According to the National Research Center on Learning Disabilities (NRCLD), an RtI model must be formally in place school-wide, to apply it to high-stakes decisions, such as special education eligibility. Further, Shapiro & Clemens (2009) suggested that successfully implementing a complex model, such as RtI, may take between three to five years indicating that implementation of an RtI service delivery model cannot be completed quickly. Rather, implementation is a thoughtful, long-term process in which schools engage in across multiple school years.

RtI Implementation. A majority of the available studies evaluating RtI implementation are qualitative and describe the practices states and/or specific school districts are adopting. In other words, most of the data collected are based on self-report from teachers, state administrators, and other school personnel. In 2007, approximately 48 out of 51 state department directors of special education reported that their state was implementing RtI or was, at least, considering the implementation of RtI (Hoover, Baca, Wexler-Love, & Saenz, 2008; Berkeley, Bender, Peaster, & Saunders, 2009). The two elements of RtI that were found to be implemented most consistently across states were progress monitoring at all tiers and the inclusion of research-based instructional services and other student support practices (Berkeley, Bender, Peaster, & Saunders, 2009). However, variability exists in how states, districts, and schools implement RtI.

Of the states that adopted RtI before 2008, more than half, or 58% to 67%, reported using RtI models that combine standard treatment protocol and problem-solving approaches (Berkeley, Bender, Peaster, & Saunders, 2009; Hoover, Baca, Wexler-Love, & Saenz, 2008). Additionally, the two most commonly reported emphasized purposes for RtI were instructional decision-making and identifying students eligible for special education services (Hoover, Baca, Wexler-Love, & Saenz, 2008). Finally, although a majority of states indicated the use of progress monitoring as an RtI practice, Bailey (2014) found that some schools reported using measures that were not research-based and only some schools reported using collected data for systematic decision-making. Varying results also have been found regarding the effectiveness of RtI implementation based on whether researchers or school personnel implemented the model.

In a 2002 meta-analysis, Burns and Symington (2002) found that problem-solving teams led by university faculty members had a greater influence on student outcomes (i.e. time on task, task completion) and systemic outcomes (i.e. referrals to special education, percentage of referrals that are diagnosed with a disability) than did problem-solving teams led by school personnel; however, both teams had a large effect on both student and systemic outcomes. In another, more recent, meta-analysis examining four large-scale implementations of RtI, implementation led by school personnel had a large effect on systemic outcomes and a medium effect on student outcomes while implementation led by researchers had a large effect on student outcomes and a small to medium effect on systemic outcomes (Burns, Appleton, & Stehouwer, 2005). It is possible that as schools have more opportunities for training

and gain more experience implementing RtI, integrity of implementation may improve if the appropriate system-level factors and support mechanisms are available.

Some of the challenges of implementing RtI reported by teachers and school personnel include limited resources, insufficient administrative support, and limited professional development opportunities and time (Bailey, 2014). Further, schools in rural districts reported that data based decision making was the most challenging RtI component to implement (Dexter, Hughes, & Farmer, 2008) and staff training was consistently identified as an area for improvement. Berkeley, Bender, Peaster, and Saunders (2009) reported that 88% of states were conducting professional development on RtI, while other studies have found that these training efforts vary in their focus and are generally inadequate for providing school personnel with the knowledge base necessary to implement RtI in an optimal way (Hoover, Baca, Wexler-Love, & Saenz, 2008; Bailey, 2014; Martinez & Young, 2011). The challenges reported are relevant to system-level supports and RtI as a school-wide initiative that requires a supportive infrastructure.

Contemporary perspectives of RtI implementation emphasize system-level factors that promote or discourage fidelity of implementation (Stobier, 2014). When students have difficulty learning, it is regarded as a mismatch between resources and student need (Greenwood & Kim, 2012; Burns & Gibbons 2008). In other words, student success is a result of creating a context that is most conducive to each student's individual learning needs. Within an RtI framework, assessment and decision-making extend beyond students' skills to the environment in which learning is occurring. Assessing the environment in addition to student skills establishes the

optimal conditions in which student learning takes place to maximize the effectiveness of instruction. When the appropriate and necessary supports are in place, a more effective RtI implementation is promoted (Noltemeyer, Boone, Sansosti, 2014). Some system-level factors that have been found to promote implementation include leadership, coordination of assessment and data management, culture and beliefs, resource allocation, and opportunities for professional development (O'Connor & Freeman, 2012; Kratochwill, Volpiansky, Clements & Ball, 2007).

Fidelity of RtI Implementation: Given the complexity of the RtI multifaceted framework, fidelity of implementation is an essential component to the implementation process. Data regarding fidelity of implementation make it possible to determine if changes in student performance are the result of the instruction and/or the intervention and further, when a change in instruction or intervention is necessary for student success. Additionally, these data make it possible to attribute student performance to RtI implementation. A majority of state department directors of special education (i.e., 36 out of 50) reported that fidelity of implementation was important to consider when implementing RtI (Berkeley, Bender, Peaster, & Saunders, 2009). Regardless, data supporting the fidelity of implementation is rarely reported in studies examining RtI implementation (Burns, Appleton, & Stehouwer, 2005; VanDerHyden, Witt, & Gilbertson, 2007). Fidelity of implementation data provide clear information that could eliminate variables related to team process and poor instruction as reasons for student failure (Fuchs, 2007). Consequently, evaluating and monitoring the RtI process would increase the credibility of the decision making process and increase confidence in decision making (Noell & Gansle, 2006). In other words, to have a

defensible RtI data set for making decisions about student learning and making decisions about student eligibility for special services, the technical adequacy of the process must be addressed.

IDEIA (2004) federal regulations allow state and local education agencies to determine and adopt varying models of RtI on an individualized basis. As a result, the implementation of RtI varies across states and districts. The number of tiers, the personnel in charge of implementation, the model for developing interventions (i.e., problem solving or standard treatment protocol), and the training opportunities are examples of elements that vary between RtI models (Dexter, Hughes, & Farmer, 2008; Berkeley Bender, Peaster, & Saunders 2009).

RtI Essential Components: According to a study surveying school principals' perceptions of RtI implementation, the variability between models has led to confusion about what specific practices or procedures must be in place to ensure that a comprehensive RtI model is implemented (Swindlehurst, Shepard, Salembier, & Hurley, 2015). Most of the research on RtI has focused on individual components of the model including tier II interventions (Fuchs, Fuchs, Mathes & Simmons 1997), decision rules for progress monitoring (Ardoin, Christ, Morena, Cormier & Klingbeil, 2013), evidence-based assessments (Deno, 1985; Stecker, Fuchs, & Fuchs 2005) and the problem solving team process (Newton et al., 2012). Studies regarding the individual components of RtI are critical to the general understanding of RtI; however, it is equally important to study RtI as a system of coordinated components and the effect of RtI on student outcomes and systemic outcomes. For example, Hill, King, Lemons and Partanen (2012) reviewed 22 studies examining the efficacy of tier II

interventions and found that most studies did not explicitly target the same skills or prerequisite skills as the tier I core instruction and, additionally, most studies did not report on the effectiveness of the tier I core instructional program. Relating the tier II intervention to the tier I core curriculum is necessary to evaluate the efficacy of the tier II intervention within the context of the multi-tiered RtI framework (i.e., if the intervention is an effective supplement to tier I core curriculum or if tier I core curriculum is appropriate and effective). Otherwise, the study is decontextualized from the broader RtI efforts and it is unclear if the results can be generalized to other schools. Currently, there is a lack of empirical validation of RtI as a complete model, which incorporates the components of RtI implemented as a coherent system.

In recent years, efforts have been made to standardize the evaluation of RtI implementation, and allow researchers to compare effects between models. Unlike many elements of an RtI framework that may vary in implementation (i.e. number of tiers, personnel in charge of implementation), these essential components must be implemented for a model to be definitively identified as an RtI model. According to the National Center for Response to Intervention (NCRTI), comprehensive RtI implementation can be distinguished by five essential service delivery components identified based on decades of research: 1) assessments, 2) data based decision making, 3) multilevel instruction, 4) infrastructure and support mechanisms, and 5) fidelity and evaluation (NCRTI, 2015). Establishing essential components allows schools to assess to what extent RtI is being implemented and how well RtI is being implemented. Most importantly, schools can determine if RtI implementation impacts student outcomes.

Relationship Between Implementation and Outcomes: In a review of studies examining RtI implementation in various rural school districts, Dexter, Hughes, and Farmer (2008) found that in general, RtI models improved student performance on reading measures and math measures. Although gains in reading and math performance were noted, these studies ranging from 1999-2007 did not provide fidelity of implementation data and as a result, improvements cannot be attributed, with confidence, to the implemented RtI model. Further, mixed results were found regarding the impact on special education placement rates ranging from rates staying constant to rates falling by seven percent (Bollman, Silberglitt, & Gibbons, 2007; Dexter, Hughes, & Farmer, 2008; VanDerHyden Witt, & Gilbertson, 2007).

VanDerHyden, Witt, and Gilbertson (2007) examined the effects of one RtI model, STEEP, within one school district on systemic outcomes including student eligibility for special services, the use of STEEP data for decision making, identification rates by ethnicity and sex, the placement costs for the district and the accuracy of decisions regarding students' response to intervention. Results indicated that using the STEEP model led to more accurate identification in which students should be tested for eligibility purposes, which reduced the number of cases referred for eligibility testing, and therefore, the district assessment costs. Additionally, this study provided initial evidence that RtI models may have the potential to decrease the disproportionality in identification rates by ethnicity and by gender.

In another study, Bailey (2014) described how seven rural schools were implementing the essential components of RtI, as defined by the NCRTI (2014). Three elementary schools, one middle school, one high school, and one Kindergarten

through twelfth grade school were included in the study, and the number of years the schools engaged in RtI implementation ranged from two to eight years. All schools engaged in some type of screening process, yet several schools indicated the data were not used for decision-making. Similarly, 74% of schools implemented a progress monitoring system; however, only 53% of those schools used the data for instructional decision-making. Therefore, it was not surprising that most schools reported data based decision making as the most difficult component to implement, and only half of the schools had teams that met regularly to discuss student data. Finally, most schools focused their efforts on tier II interventions and only one school reported providing evidence-based tier III supports. One limitation of this study is how information was summarized across school levels that have been implementing RtI for a large range of years.

Another study (Noltemeyer & Sansosti, 2012) sought to examine the association between school-level RtI implementation (i.e. Ohio's Integrated Systems Model, which incorporates behavior into the typical academic RtI model) and student reading outcomes, measured by DIBELS oral reading fluency measures. An adherence checklist measured the percentage of RtI academic features and behavior features that were in place within the school. Results of the study indicated that academic implementation of RtI significantly predicted student performance on DIBELS oral reading fluency measure above and beyond controlling for student's initial scores on the oral reading fluency measure from fall of the prior school year.

Finally, Sharp, Sanders, Noltemeyer, Hoffman, and Boone (2016) conducted a study to examine the relationship of school-level RtI implementation and reading

achievement. An empirically validated self-report RtI measurement scale, called the RtI Implementation Scale for Reading (RTIS-R), was used, which was developed based on five essential components of RtI (i.e. assessment, data based decision making, instruction, professional development, and treatment integrity) (Noltemeyer, Boone, & Sansosti, 2014). Results revealed three significant models that predicted student scores on the state achievement test. In the first model, the percentage of economically disadvantaged students explained 27.8% of the variability in scores and in the second model, the number of discipline referrals explained 8.1% of variability in scores. Finally, in the third model, data based decision making explained 7.2% of variability among scores above and beyond the demographic variables. Although the external validity of this study is limited, it is the first to quantitatively examine the relationship between the quality of school-level RtI implementation as a coherent system, and student level reading achievement. In addition, these results suggest that RtI implementation positively contributes to students' reading performance even after controlling for demographic variables.

Purpose of the Study

General knowledge and understanding of RtI educational service delivery methods and models is increasing as a result of a growing research base. However, there are several limitations to the available research on evaluating the fidelity of RtI implementation. First, the number of studies quantitatively examining RtI implementation as a complete, coherent system appears sparse (Glover & DiPerna, 2007). Second, due to the inconsistent definition of RtI, some studies may report to be examining the large-scale implementation of RtI, but may in fact be examining only a

small-scale implementation (i.e. only some essential components are implemented) (Burns & Symington, 2002). If RtI implementation is defined by five essential service delivery components, then comparisons cannot be made between models with all five components implemented and models with less than five components implemented. Finally, from the limited studies available, mixed results have been found regarding whether RtI implementation predicts student outcomes, or not.

The purpose of the current study is to extend the extant literature by using a rubric, developed by the National Center on Response to Intervention, to quantitatively evaluate RtI implementation and its' relationship to student reading outcomes in one large education district with a long-standing history of RtI implementation. Exploring how schools are implementing RtI, including the relationships between implementation and student outcomes, will help researchers and practitioners gain a better understanding of the pathway to optimal RtI implementation, which may vary based on school characteristics. Specifically the following research questions were addressed:

1. To what extent are staff in elementary schools implementing the essential components of RtI with integrity, as measured by the RtI Essential Components Worksheet between schools; and if differences exist between schools, how can those differences be characterized?
2. To what extent does the integrity of overall RtI implementation, and the implementation of the five essential components, in elementary schools, relate to

differences student-level student outcomes, at each grade across schools, as measured by AIMSWEB screening data?

Methodology

Procedures

The Research Review Committee at the education district where the data collection occurred approved the current study. The signed letter of approval was sent to the Office of Research Compliance at the University of Rhode Island, and the decision was made that the study did not require further approval for human subject research because it would be utilizing an extant data set.

Data collection occurred during the 2014-2015 school year. The original study began as a conversation between special education personnel and administrators in the education district to address the district's progress on the implementation of RtI. More specifically, these individuals were interested in learning what was working and what was not working regarding the RtI implementation process. Prior to this study, the district used self-report surveys to assess the implementation of RtI. However, administrators and special education personnel thought the results were inflated, which is a common limitation with self-report measures (Cook & Campbell, 1979). As a result, the executive director for the education district sought to find an objective measure to assess the quality of RtI implementation. In 2014, the education district chose to use measures created by the National Center on Response to Intervention as an objective measure of RtI implementation.

Following the selection of the measures, the executive director spoke to administrators from each member district to discuss the study and requested for a group of key personnel involved in RtI to be selected for a group interview. Next, the executive director set up meetings at each school with the selected teams. One two-

hour meeting was scheduled at each school during March and April of 2015 to conduct the semi-structured interview. Three individuals were involved with conducting the interviews. These individuals were highly involved with the overall implementation of RtI throughout the education district, as their positions were shared core services that were involved in all member districts. Their positions included executive director for the education district, instructional services coordinator, and director for special education services. As a result, each person was highly knowledgeable about the continued efforts and current practices in place. Further, their experience working within the district varied; one individual had over 10 years of experience working in the district, one individual had a few years of experience in the district, and one individual was new to the district the year of the data collection. During the interview, one individual facilitated the conversation and the other two individuals took notes. The note-takers also monitored the quality and quantity of responses to ensure enough information was collected to properly rate the responses on the rubric, and these individuals asked clarifying questions, as necessary. For example, clarifying questions included asking for permanent products to review. The executive director facilitated seven interviews, and the instructional services coordinator facilitated one interview.

Once the interviews were finished, the executive director and the instructional services coordinator, both trained in school psychology, separately scored the interviews using the RtI Fidelity of Implementation Rubric. Each item on the semi-structured interview was assigned a rating. Inter-rater reliability was not calculated, and the information to calculate it was unavailable. However, the individuals

compared their scoring and when disagreements were found they discussed their point of views, used their knowledge of the school, and came to a consensus on the best rating. In the end, the rater's agreed on a rating for each of the 31 items. This information was entered into an excel worksheet, school names were de-identified, and averages were calculated for each of the five components of RtI, as well as, for the total RtI score, which was the total of the five components. The data file was sent to the researcher of the current study.

Student benchmark data were collected three times during the school year: September (i.e. fall), January (i.e. winter), and April (i.e. spring). Following data collection at each time point, students' scores were entered on the AIMSweb online data management and reporting system. The outcomes services manager exported data from the AIMSweb website to an excel sheet to provide the researcher with R-CBM words read correct (WRC) scores and rate of improvement (ROI) scores for students in first grade through fifth grade at each of the eight elementary schools. Student data were de-identified using unique identification numbers to maintain student anonymity.

Measures

Response to Intervention/Instruction. The RtI Essential Components Worksheet (Appendix B) and the accompanying RtI Fidelity of Implementation Rubric (Appendix C) were used to measure the schools' quality of RtI implementation. Together, the two measurement instruments (i.e. worksheet and rubric) construct the RtI integrity framework to evaluate school-level implementation of RtI. The National Center for Response to Intervention/Instruction (NCRTI), a federally funded center, in partnership with RMC Research Corporation, developed

the RtI integrity framework as a tool to monitor and evaluate RtI implementation across schools (Elledge, 2010). More specifically, the development was based on the expertise of the NCRTI's senior advisors and center leadership, with the purpose of helping schools understand if they were implementing RtI well and what to look for when evaluating their RtI implementation. The instruments were developed for use by RtI coordinators or evaluators with extensive RtI knowledge, and who are qualified to make judgments and ratings about RtI implementation (Elledge, 2010).

The RtI Essential Components Worksheet is organized according to the five essential components of RtI, as identified by the NCRTI. Multiple items were identified to define each essential component. The worksheet contains five items related to assessments, three items related to data based decision-making, 12 items related to multi-level instruction, nine items related to infrastructure and support mechanisms, and two items related to fidelity and evaluation. All 31 items are used to establish overall RtI implementation. The worksheet was designed for use as a semi-structured interview and sample questions are provided to guide a discussion between the interviewer/s and the interviewees. In addition to the discussion, the interviewer is encouraged to conduct observations and document reviews to collect sources of evidence and a rationale for rating each item (Elledge, 2010).

Following the interview, observations, and document reviews, a rating was assigned to each item on the worksheet using the RtI Fidelity of Implementation Rubric. The rubric provides a likert-type response format for each item, and schools may be assigned a one, two, three, four or five from the rubric, with a one being the lowest quality and a five being the highest quality of implementation. A narrative

rationale is available for a rating of a one, a three, and a five. Although a narrative rationale is not available for a rating of a two and a four, it is clear that a rating of a two falls between a one and a three, and a rating of a four falls between a three and a five. After the ratings were assigned for each individual item, the mean of the items for each essential component was calculated, and that rating was used as the single value that represented each essential component. For the purposes of the current study, a mean rating of four or higher indicates high quality RtI implementation and a mean rating less than 4 indicates low quality RtI implementation.

Assessment. Educational assessments are measures used to determine what students know and are able to do before, during, and after instruction (Green & Johnson, 2010). Three types of assessments are used within an RtI framework to inform decision making including screening, progress monitoring, and other supporting assessments (i.e. curriculum-based measures, state achievement tests) (NCRTI, 2014). The purpose of assessment is to provide early identification of students who may be struggling to meet grade level expectations by monitoring all students' responses to the general curriculum, as well as, to interventions (Burns & Gibbons, 2008).

The assessment component has a total of five items on the measure, and it is broken into two categories: screening and progress monitoring. Screening is described as measures used to identify students at risk of poor learning outcomes or challenging behavior, and three items fall under screening (i.e. screening tools, universal screening, & data points to verify risk). Progress monitoring is described as ongoing and frequent monitoring of progress to quantify rates of improvement and to inform

instructional practice, as well as, the development of individualized programs. Two items fall under progress monitoring (i.e. progress-monitoring tools & progress monitoring process). As such, schools may earn a maximum of 25 points on this component. A low rating indicates that there is a need for improvement in assessment practices, and a rating of a five indicates that all conditions included in the item are being implemented. For example, on the item ‘data points to verify risk’, a rating of a one is described as, “Screening data are not used or are used alone to verify decisions about whether a student is or is not at risk,” and a rating of a five is described as, “Screening data are used in concert with at least two other data sources to verify decisions about whether a student is or is not at risk (NCRTI, 2014).”

Data-based decision making (DBDM). DBDM represents the processes used to inform instruction, intensity of intervention (i.e. movement within multilevel tiers), and disability identification (NCRTI, 2014). The DBDM component has a total of three items, which include decision making process, data system, and responsiveness to secondary and intensive levels of intervention. Schools may earn a maximum of 15 points on the DBDM component. A low rating indicates that there is a need for improvement in DBDM practices, and a rating of a five indicates that all conditions of DBDM are being implemented. For example, on the item ‘data system,’ a rating of one is described as, “A data system is in place that meets two or fewer of the following conditions, 1) the system allows users to document and access individual student-level data and instructional decisions; 2) data are entered in a timely manner; 3) data can be represented graphically; and 4) there is a process for setting/evaluating

goals,” and a rating of five is described as a data system that meets all four conditions (NCRTI, 2014).

Multi-tiered systems for support (MTSS). MTSS refers to a multilevel system of instruction and intervention for preventing student failure, and it is commonly characterized by a three-tiered triangle. The MTSS component has a total of 12 items and schools may earn a maximum of 60 points on the MTSS component. MTSS is broken down into three categories: Tier I, Tier II, and Tier III. Five of the items fall under the category of primary level instruction/core curriculum, or Tier I (i.e. research-based curriculum materials, articulation of teaching and learning, differentiated instruction, standards-based, & exceeding benchmark), four items fall under secondary-level intervention, or Tier II (i.e. evidence-based instruction, complements core instruction, instructional characteristics, & addition to primary), and the remaining three items fall under intensive intervention, or Tier III (i.e. data-based interventions adapted based on student need, instructional characteristics, & relationship to primary). A rating of a one is described as meeting none or one of the necessary conditions, and a rating of a five suggests that all conditions of MTSS are met. For example, on the item ‘research-based curriculum materials’ a rating of one is described as, “Few core curriculum materials are research based for the target population of learners,” and a rating of five is described as, “All core curriculum materials are research based for the target population (NCRTI, 2014).”

Infrastructure and support mechanisms. Infrastructure and support mechanisms represent the knowledge, resources, and organizational structures that are necessary to clearly define all components of RtI in a unified system to meet the

established goals (NCRTI, 2014). The infrastructure and support component has a total of nine items; however, two items on this component were not scored for the current sample of schools. As a result, these items were not included and a total of seven items comprised the component, with schools able to earn a maximum of 35 points (i.e. prevention focus, leadership personnel, school-based professional development, schedules, resources, cultural & linguistic responsiveness, communications with & involvement of parents, communication with & involvement of staff, and RtI teams). A low rating suggests that improved supports are needed to effectively implement RtI and a high rating indicates that appropriate supports are in place and available for RtI implementation. For example, on the item, ‘resources,’ a rating of one is described as, resources are not allocated to support implementation, a three is described as resources are partially allocated, and a five indicates that resources are adequately allocated for implementation (NCRTI, 2014).

Fidelity and evaluation. Fidelity and evaluation is described as a system for collecting and analyzing data to measure the adherence to and effectiveness of the RtI model (NCRTI, 2014). Two items are incorporated in this component, and schools may earn a maximum of 10 points (i.e. fidelity & evaluation). On this component, a rating of a one suggests that none of the conditions are met; a rating of three suggests that at least one condition is met, and a rating of five suggests that all conditions are met. For example, on the item ‘fidelity,’ a rating of a one is described as “Neither of the following conditions is met: 1) procedures are in place to monitor the fidelity of implementation of the core curriculum and secondary and intensive interventions” and a rating of a five is described as “Both of the conditions are met (NCRTI, 2014).”

Measure validity and reliability. The RtI fidelity of implementation worksheet and rubric are valid measures of RtI. The instruments' content validity was established by thorough and intensive review and evaluation by experts in the field of RtI (i.e. Doug and Lynn Fuchs, Sharon Vaughn, Don Deschler, Lou Danielson, and Stephanie Jackson). These experts are highly regarded leaders in the field of special education, who are well known for their long-standing involvement in researching RtI. Additionally, these experts worked closely to ensure the appropriateness, breadth, and clarity of the content (NCRTI, personal communication, November 19, 2015). In other words, the items included on the measures are representative of the overall concept of an RtI framework. The original version of the measures were released in 2011, and since, there have been two revisions (i.e. 2014 and 2015) to reflect recent research findings, experience, and language changes in the field, as well as, changes to simplify the structure (NCRTI, personal communication, November 19, 2015). Additionally, the measures have face validity, or appear to be a reasonable measure of RtI, which is evident by the use of the measures in school districts across the nation. Although these measures have been used in prior studies, information regarding the internal consistency of the measure was not available (Bailey, 2014; Maier, Pate, Gibson, Hilgert, Hull, & Campbell, 2016). In the current study, the internal consistency coefficient for the RtI measure was $r_{\alpha} = .776$, indicating acceptable reliability of the measure (Nimon, Zientek, Henson, 2012). Further, the internal consistency suggests that the set of 31 items measures a single underlying dimension, RtI. Appendix D presents a detailed discussion of the scale reliability.

Student Outcomes

Oral Reading Fluency (ORF). AIMSweb oral reading fluency (R-CBM)

measures are standardized, curriculum based assessments that are individually administered to all students three times a year. R-CBM is a measure of students' ability to read fluently with connected text (Good & Jefferson, 1998). A trained examiner administers the assessment by having a student read aloud for one minute from three grade-level passages, and the median number of words read correctly aloud per minute from the three passages is recorded. On this measure, student performance is measured by words read correct (WRC) in one minute. The same three passages are administered three times (i.e. fall, winter, spring) during the school year to screen for students who may or may not meet grade-level benchmarks, a process known as screening. National norms are available as a criterion for evaluating student success compared to other similar grade, or age, students across the nation. However, local norms were developed specific to the state in which the data were collected.

Development of local norms was based on the predictability of R-CBM scores to a state achievement test. The education district chose R-CBM as a student reading outcome measure because it is empirically validated as an indicator of a broader set of literacy skills (Good & Jefferson, 1998). Further, the results from the R-CBM are used in the education district to evaluate the effectiveness of core instruction and for data based decision making purposes including resource allocation and levels of support for students.

Rate of Improvement (ROI). Rate of improvement (ROI) is a measure of how raw scores on the R-CBM increase during a given school year. Screening data may be

used to calculate a ROI in terms of the number of words read correctly per minute gained per week. To calculate the ROI, the difference between the first and last scores is divided by the number of weeks in between collecting the scores (Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993). As a result, using screening data, a ROI may be calculated for fall to winter, for winter to spring, and for fall to spring. Students' ROI informs schools if students are making adequate progress to meet end-of-year benchmark goals.

Participants

School Characteristics. This study was a preliminary study for a larger project that will examine district-level implementation of RtI in a rural education district and its' relationship to student outcomes. The project was conducted within an education district in the Midwest region of the United States during the 2014-2015 school year. For the present purposes, an education district is defined as a specialized type of union school district which affords its' member districts certain incentives if they create a union school district. More specifically, incentives may include fiscal equity, shared programs and core services (i.e. an executive director of the education district, a director of special education, an instructional services coordinator, services for students with physical and other health impairments), and administrative effectiveness that would otherwise be unaffordable by the member districts. In other words, a governing board, formed by representatives from each member district, makes decisions regarding funding and educational services.

Figure 1. Visual representation of the structure of the education district.

The education district involved in the present project is composed of six member districts, with a total of eight elementary schools. Each district has one elementary school with the exception of one district, which has three elementary schools. Two of the three elementary schools are linked because students who attend one for kindergarten through second grade are filtered into the other for third through fifth grade. Figure 1 provides a visual representation of the education district. To maintain anonymity of the schools and students, pseudonyms were used. As a result, eight elementary schools participated in the study. Table 1 provides a summary of the participating schools' characteristics. Information about the schools' characteristics and demographics were obtained from publicly available information obtained through the state department of education's website for the 2014-2015 school year.

Overall, the education district served approximately 4,605 students in their eight elementary schools during the 2014-2015 school year. All of the elementary schools were public schools and three of the schools were Title I schools. Additionally, five of

the eight schools met or exceeded the expected achievement growth for adequate yearly progress (AYP) in reading in 2014, which is measured by state achievement tests. Further, the member districts are similar in that the same governing board makes decisions regarding resource allocation and educational support services for students. As such, each school had similar access to instructional support services and outcome management services. Moreover, each school administered AIMSweb's Reading Curriculum Based Measure (R-CBM) probes three times annually (i.e. September, January, and April) to all students as a benchmark measure for monitoring all students' progress.

Table 1
Demographic Characteristics by School

School	Grades	Population	Title Status	AYP – Reading 2014	% Special Education	% Nonwhite Students	% Free/Reduced Lunch
1	PreK-5	373	Title I	Y	14.2%	5.9%	33.5%
2	K-2	485	School	Y	8%	4.9%	23.9%
3	3-5	546	School	Y	12.6%	5.1%	24.7%
4	PreK-6	397	Title I	N	13.1%	14.9%	61%
5	PreK-6	480	Title I	N	15.4%	30%	64.6%
6	K-6	1,008	School	Y	11.5%	6.6%	34.4%
7	K-6	846	School	N	10.4%	6.6%	47.9%
8	K-6	470	School	Y	11.7%	6.6%	34%

School 1 is a Title I (i.e. receives financial assistance through federal grants available to districts serving low-income students) pre-kindergarten through fifth grade elementary school serving 373 students. In school 1, over a quarter (i.e. 33.5%) of students received free or reduced lunch, 5.9% were nonwhite students, and 14.2% of students received special education services.

School 2 is a kindergarten through second grade elementary school that served 485 students. In school 2, approximately a quarter (i.e. 24%) of the students received

free or reduced lunch, 4.9% were nonwhite students, and 8% of students received special education services. Students who attended school 2 were filtered into school 3 for grades three through five.

School 3 is a third through fifth grade elementary school serving 546 students. In school 3, a quarter of the students (i.e. 24.7%) received free or reduced lunch, 5.1% were nonwhite students, and 12.6% of students received special education services.

School 4 is a Title I pre-kindergarten through sixth grade elementary school serving 397 students. Over half of the students (i.e. 61%) received free or reduced lunch, 14.9% were nonwhite students, and 13.1% of students received special education students.

School 5 is a Title I pre-kindergarten through sixth grade elementary school serving 480 students. Over half of the students (i.e. 64.6%) received free or reduced lunch, 30% were nonwhite students, and 15.4% of students received special education services. School five had the greatest diversity of all the elementary schools in the education district.

School 6 is a kindergarten through sixth grade elementary school serving 1,008 students. In school 6, 34.3% of students received free or reduced lunch, 6.6% were nonwhite students, and 11.5% of students received special education students.

School 7 is a kindergarten through sixth grade elementary school serving 846 students. Approximately half (i.e. 47.9%) of the students received free or reduced lunch, 6.6% were nonwhite students, and 10.4% of students received special education services.

School 8 is a kindergarten through sixth grade elementary school serving 470 students. Over a quarter of the students (i.e. 34%) received free or reduced lunch, 6.6% of students were nonwhite students, and 11.7% of students received special education services.

Interviewees. A small group, of approximately four to eight individuals, from each school participated in the semi-structured interview. With respect to the selection of individuals, faculty and/or staff who were identified as the key people involved with RtI implementation were invited, directed, or asked to attend by administrators in the district. Nellis (2012) discussed the variability in key school personnel involved with RtI implementation between schools. Individuals were invited to participate; they were not required to participate. Each school had different types of personnel serving on the interview team. All teams included an administrator, school psychologist, guidance counselor and intervention teacher. Additionally, teams varied in their inclusion of a special education teacher, general education representative, data specialist and literacy specialist. A total of four to eight individuals were chosen to represent each school. Specific information regarding the demographic characteristics of these individuals was not available.

Students. Students with benchmark scores available at all three time points (fall, winter, spring) were included in the study. Approximately 100 students were removed from the study due to missing data. The final sample consisted of 2,901 students attending eight elementary schools. See Table 2 for the total number of participating students by school and by grade.

Schools 4, 5, 7, and 8 had student data available at all five grade levels. School 4 had a total of 269 students participants, school 5 had a total of 383 student participants, school 7 had a total of 591 student participants, and school 8 had a total of 319 student participants. School 6 had the most (i.e. 606) student participants and student data were available for grades one, two, and three. School 1 had a total of 241 student participants and student data were available for grades one, two, three, and four. School 2 had the least amount of student data available (i.e. 241 student participants), as it was the smallest school serving only grade one and grade two. Finally, school 3 had 351 student participants and student data were available for grades three and four.

Table 2
Total Number of Participants by School and Grade

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Total students per school
School 1	60	69	48	64	N/A	241
School 2	141	N/A	N/A	N/A	N/A	141
School 3	N/A	N/A	177	174	N/A	351
School 4	58	52	59	58	42	269
School 5	83	83	72	70	75	383
School 6	190	202	214	N/A	N/A	606
School 7	113	120	119	122	117	591
School 8	54	65	74	65	61	319
Total students per grade	699	591	763	553	295	2901

Informed Consent

Informed consent was not required for the current study. AIMSweb screening data were collected as part of a prevention framework that all students participate in as part of federal regulations for schools to identify all children who may require special education services (IDEIA, 2004). As a result, schools are permitted to administer

these measures without collecting informed consent from parents. Further, school level data and student level data were de-identified before sent to the researcher.

Design

The current study employed a quasi-experimental, non-equivalent groups design. This design is commonly used in applied educational research because students cannot be randomly assigned into groups. Rather, intact groups based on student enrollment at a particular school are used. As a result, the groups are not equivalent, or as similar as they would have been if random assignment were used. It is important to note because any prior differences between the groups may affect the study outcomes and conclusions (Cook & Campbell, 1976).

Results

Analyses

RtI integrity of implementation data were available for a total of eight elementary schools. The first part of the current study sought to examine differences in the fidelity of RtI implementation between the eight elementary schools. As a result of the small sample size (N=8) and the limited variability in the integrity of RtI implementation between schools, questions one and two were integrated, and examined using descriptive statistics (i.e. means, standard deviations, bar charts).

The second part of the current study sought to examine the relationship between fidelity of RtI implementation and overall student reading outcomes by school and by grade level. Student outcome data were not available for all grade levels at each school; thus, student outcomes were analyzed by grade level. Further, based on the available school-level data from the extant data set, and the small sample size (N=8) it was concluded that prediction methods were not the most appropriate method for analysis. Consequently, the third and fourth research questions were combined and modified to reflect this change. The reader is referred to Appendix A for a further explanation of the edited research questions. Rather, the new, edited, research question was analyzed using analysis of variance (ANOVA) to examine differences in student outcomes between schools, for each grade level. As such, 10 separate one-way ANOVA's were conducted. The relationship between fidelity of RtI implementation and student reading outcomes was analyzed qualitatively.

First Research Question

Descriptive statistics. The first research question explored to what extent staff in elementary schools are implementing the essential components of RtI with integrity, as measured by the RtI Essential Components Worksheet, between schools; and if differences exist between schools, how those differences can be characterized. A total of eight teams, one from each elementary school, were interviewed and assigned ratings on the RtI Fidelity of Implementation Rubric. As such, eight mean scores were available for comparison on each of the five RtI components and for the total RtI score. Table 3 represents mean scores, standard deviation, range of scores, minimum score earned, and the maximum score earned for each of the five components of RtI and the total RtI score. Out of 145 possible points, the total RtI scores ranged from 102 points (i.e. school 8) to 133 points (i.e. school 2). In regard to the five essential components, assessment received the highest mean score across schools ($M= 4.875$, $SD=.1488$) followed by data based decision making ($M= 4.6250$, $SD= .2799$). Infrastructure and evaluation received the lowest mean score ($M=3.3738$, $SD= .59783$) across schools. In addition to mean scores, range of obtained scores on each component is reported. These scores were derived by calculating the difference between the highest and lowest values obtained on each component, which provides a comparison of the variability in scores between schools for each component. Assessment was the most consistently implemented component across schools (i.e. range = .4), followed by data based decision making (i.e. range= .67). The MTSS and fidelity and evaluation components both had a range equal to 1.5. The greatest range of mean ratings was observed on the infrastructure and support component (range=

1.58). At least one school received a perfect score of 5 on the assessment component and on the data based decision making component. For the remaining components, the maximum score received by a school was 4.58 on MTSS, 4.29 on infrastructure and support, and 4.0 on fidelity and evaluation. Each component is discussed in more detail below. For the purposes of the current study, ratings of 1, 2, and 3 are indicative of lower quality implementation and ratings of 4 and 5 are indicative of higher quality implementation.

Table 3

Descriptive Statistics for the RtI Measure

	N	Mean	SD	Range	Minimum	Maximum
Assessment	8	4.875	.1488	.4	4.6	5.00
DBDM	8	4.6250	.27990	.67	4.33	5.00
MTSS	8	3.7188	.46218	1.50	3.08	4.58
Infrastructure & Support	8	3.3738	.59783	1.58	2.71	4.29
Fidelity & Evaluation	8	3.375	.6409	1.5	2.5	4.0
Total RtI	8	114.75	9.910	31	102	133

RtI total. Out of 145 possible points to earn, the eight elementary schools earned between 102 points (i.e. school 8) and 133 points (i.e. school 2). A score greater than or equal to 116 points indicates high quality implementation (i.e. a minimum rating of 4 on each of the 29 items for which ratings are available). Three schools fell above the cut-off for high implementation, with one school earning 133 points (i.e. school 2) and two schools earning 121 points (i.e. schools 1 & 3). The total RtI score in the remaining elementary schools was categorized as low quality implementation. One elementary school earned 114 points (i.e. schools 4) and three

elementary schools earned less than 110 points (i.e. schools 5, 6 & 8). Table 4 presents a comparison of all ratings on the RtI measure across schools.

Table 4

<i>Ratings on the RtI measure by school</i>						
School	Mean Assessment	Mean DBDM	Mean MTSS	Mean Infrastructure & Support	Mean Fidelity & Evaluation	Total RtI
1	5*	4.67*	4*	3.86	3.5	121*
2	5*	5*	4.58*	4.29*	4*	133*
3	5*	4.67*	4*	3.71	4*	121*
4	5*	4.67*	3.67	2.86	4*	114
5	4.8*	4.33*	3.5	2.71	2.5	107
6	4.6*	4.33*	3.5	2.71	3.5	108
7	4.8*	5*	3.42	3.71	3	112
8	4.8*	4.33*	3.08	3.14	2.5	102

*Meets the criterion for high implementation

Assessment. Figure 2 presents mean assessment ratings for each of the eight elementary schools. All of the elementary schools received a mean rating greater than 4.5, demonstrating a high level of quality and similarity of assessment implementation. Four of the elementary schools received the highest possible score, a mean rating equal to 5 (i.e. schools 1, 2, 3, and 4). Three of the elementary schools received a mean rating equal to 4.8 (i.e. schools 5, 7, and 8) and one elementary school received a mean rating equal to 4.6 (i.e. school 6), which was the lowest mean rating obtained for assessment. The assessment component had the smallest amount of overall mean differences between schools, indicating it is the most consistent RtI component implemented across schools. Further, the greatest difference in means between the schools was .4, which does not appear to be meaningful. Overall all eight schools demonstrated high quality assessment implementation.

Figure 2. Mean assessment ratings by school.

Data based decision-making (DBDM). Figure 3 presents mean DBDM ratings for each of the eight elementary schools. All eight elementary schools received a mean rating greater than 4, demonstrating high quality of implementation across schools. Two elementary schools received a mean rating equal to 5 (i.e. schools 2 & 7), three elementary schools received a mean rating equal to 4.67 (i.e. schools 1, 3, & 4), and three elementary schools received a mean rating equal to 4.33 (i.e. schools 5, 6 & 8). The greatest difference observed between elementary schools was equal to .77, which does not appear to be meaningful. However, these data indicate that two of the schools are implementing data based decision making at the highest level, while five schools continue to improve their implementation efforts.

Figure 3. Mean DBDM ratings by school.

Multi-tiered systems of support (MTSS). Figure 4 presents mean MTSS ratings for each of the eight elementary schools. All eight elementary schools received a mean rating greater than 3. Ratings on the MTSS component ranged from 3.08 to 4.58, indicating greater variability on this component than for the assessment and the data based decision-making components. Further, six different mean scores resulted, indicating six levels of implementation across the eight elementary schools. Three schools received a mean rating greater than or equal to 4, indicating high quality implementation (i.e. school 2 $M=4.58$, school 1 $M=4$, & school 3 $M=4$). The remaining schools similarly fell between a mean rating of 3 and 4, indicating lower quality of implementation compared to the other schools. More specifically, school 4 received a mean rating of 3.67, schools 5 and 6 received a mean rating equal to 3.5, school 7 received a mean rating equal to 3.42, and finally, school 8 received a mean rating of 3.08. The greatest difference (i.e. 1.5) observed was between school 2 and school 8.

Figure 4. Mean MTSS ratings by school.

Infrastructure and support. Figure 5 presents mean infrastructure and support ratings for each of the eight elementary schools. All schools received a mean

rating higher than a 2.5. Mean ratings on the infrastructure and support component ranged from 3 to 4.29. Only one school received a mean rating greater than or equal to 4 (i.e. school 2, $M= 4.29$), indicating higher quality of implementation. The remaining schools received mean ratings less than 4, indicating lower quality implementation. Three schools received ratings between 3.5 and 4 (i.e. school 1, 3, & 7), one school earned a rating between 3 and 3.5 (i.e. schools 8), and the remaining three schools earned ratings between 2.5 and 3 (i.e. 4, 5 & 6). The greatest difference was observed between school 2 and schools 5 and 6 (i.e. 1.58). Schools 5 and 6 both earned the same mean rating (i.e. 2.71), indicating similarities in their quality of implementation. Finally, given that only one school fell in the high implementation range, these results indicate that infrastructure and support is an area that generally needs improvement.

Figure 5. Mean infrastructure and support ratings by school.

Fidelity and evaluation. Figure 6 presents mean fidelity and evaluation ratings for each of the eight elementary schools. Mean ratings on the fidelity and evaluation component ranged from 2.5 to 4, indicating a greater range than the assessment and DBDM components, and a similar range to the MTSS and infrastructure and support components. However, fidelity and evaluation ratings had

the greatest variation from the mean score. Five schools received their lowest RtI score in the area of fidelity and evaluation (i.e. schools 1, 2, 5, 7, & 8). Three elementary schools received a mean rating equal to 4 (i.e. schools 2, 3, & 4), indicating a high level of implementation. The remaining five elementary schools received ratings less than 4, indicating lower quality implementation: two elementary schools received a mean rating equal to 3.5 (i.e. schools 1 & 6), one elementary school received a mean rating equal to 3 (i.e. school 7), and two elementary schools received a mean rating of 2.5 (i.e. schools 5 & 8). Consequently, the greatest difference was observed between the schools with a mean rating of 4 (i.e. schools 2, 3 & 4) and schools with a mean rating of 2.5 (i.e. schools 5 & 8). Overall the fidelity and evaluation component received the lowest mean ratings and indicated the area for greatest improvement.

Figure 6. Mean fidelity and evaluation ratings by school.

Summary. The data reported here suggests that there are similarities and differences in the extent to which staff in elementary schools are implementing the essential components of RtI with fidelity. All eight elementary schools received ratings in the high implementation range in the areas of assessment and DBDM,

indicating similar, high quality practices (i.e. greater than or equal to 4.33).

Alternatively, differences in RtI implementation are characterized by MTSS, infrastructure and support, fidelity and evaluation, and, consequently, total RtI score. The greatest difference between the highest and the lowest ratings between schools was observed for infrastructure and support, followed closely by MTSS and fidelity and evaluation. Further, ratings were the least consistent on the fidelity and evaluation and the infrastructure and support components.

In regard to the quality of RtI implementation, school 2 had the highest level of implementation, earning ratings in the high implementation range on total RtI and all five components. School 3 had the second highest level of implementation, earning ratings in the high implementation range for total RtI and four components (i.e. assessment, DBDM, MTSS, & fidelity & evaluation). Next, school 1 earned ratings in the high implementation range for total RtI and three components (i.e. assessment, DBDM, & MTSS). Then, school 1 earned ratings in the high implementation range on three components (i.e. assessment, DBDM, & MTSS). Finally, schools 5, 6, 7, and 8 earned ratings in the high implementation range on two components (i.e. assessment & DBDM).

Second Research Question

Descriptive statistics and preliminary analyses. The second research question examined to what extent integrity of RtI implementation relates to differences in student outcomes, at each grade level across schools, as measured by AIMSweb oral reading fluency (ORF) data. Descriptive statistics include sample size, mean, and standard deviation are presented below for both words read correct per minute (WRC)

(i.e. Table 5) and rate of improvement (ROI) (i.e. Table 6). Separate analyses were conducted for each grade level because data were not available for all grades at each school. Multiple methods were used to assess the normality of the distribution for both student outcome variables at each grade level. First, skewness and kurtosis were examined according to guidelines presented by Harlow (2014). Next, visual representations, including histograms, normal Q-Q plots, and boxplots were reviewed to verify normality of the data and to identify outliers. Finally, the Shapiro-Wilk statistic was obtained and a non-significant result indicated normality of the distribution. Five variables were found to have a non-normal distribution. For these variables, the nonparametric alternative to ANOVA (i.e. Kruskal-Wallis test) was conducted. For three of the variables (i.e. Grade 1 ROI, Grade 3 ROI, & Grade 5 ROI), results from the nonparametric test were in agreement with the results from the ANOVA, and the researcher chose to report the results from the ANOVA. For two variables (i.e. Grade 1 WRC & Grade 4 ROI) the results from the nonparametric test differed from the ANOVA results, and the researcher chose to report the results from the nonparametric equivalent test. Results from the normality testing are discussed below. A detailed comparison of parametric and nonparametric tests is presented in Appendix E. In addition to normality, homogeneity of variances is an assumption of an analysis of variance. The Levene test was conducted and reported below for each ANOVA to assess equal variances across samples.

Table 5

Mean and Standard Deviations on the Measure of Spring WRC by School and Grade Level

School	Grade				
	1 (N= 699)	2 (N= 590)	3 (N= 762)	4 (N= 553)	5 (N= 294)
1	75.85 (32.862)	111.16 (34.556)	142.10 (45.34)	154.27 (48.666)	N/A
2	81.5 (38.084)	N/A	N/A	N/A	N/A
3	N/A	N/A	133.8 (39.018)	153.28 (42.408)	N/A
4	86.52 (42.003)	108.33 (39.457)	133.53 (43.293)	139.31 (48.059)	149.07 (45.347)
5	69.67 (31.605)	96.07 (33.891)	120.33 (44.668)	132.13 (43.514)	143.13 (40.86)
6	72.67 (35.616)	110.36 (35.819)	130.09 (41.878)	N/A	N/A
7	78.57 (28.113)	103.86 (37.585)	139.52 (40.746)	140.70 (44.086)	160.4 (44.655)
8	75.15 (36.677)	101.95 (34.479)	139.61 (36.67)	144.77 (38.209)	170.49 (39.864)

Table 6

Mean and Standard Deviations on the Measure of ROI by School and Grade Level

School	Grade				
	1 (N= 698)	2 (N= 585)	3 (N= 758)	4 (N= 547)	5 (N= 291)
1	2.0185 (.90629)	1.5725 (.41975)	1.3898 (.56196)	1.2702 (.48082)	N/A
2	1.7087 (.79696)	N/A	N/A	N/A	N/A
3	N/A	N/A	1.3025 (.55456)	1.0808 (.50841)	N/A
4	2.1174 (1.0201)	1.3012 (.48628)	1.5117 (.62316)	1.03 (.63482)	1.021 (.48828)
5	1.3502 (.84866)	1.1811 (.42907)	1.1357 (.54690)	1.0217 (.54547)	.8124 (.45146)
6	1.6353 (.95659)	1.4376 (.47475)	1.2278 (.50286)	N/A	N/A
7	1.9054 (.81016)	1.3689 (.41791)	1.4410 (.56495)	.9257 (.44415)	1.0663 (.52163)
8	1.9152 (1.0175)	1.3387 (.36385)	1.4599 (.70227)	1.108 (.48199)	1.0357 (.38266)

Grade one.

Grade one data were available for seven schools (i.e. school 1, 2, 4, 5, 6, 7, & 8).

Spring words read correct. The sample consisted of 699 first grade students.

The test for normality, examining standardized skewness and kurtosis, indicated the data were statistically normal. However, further examination, including the Shapiro-Wilk and examination of Q-Q plots tests, revealed that the data did not meet the assumption of normality due to outliers in the data. As a result, a Kruskal-Wallis test was conducted. Results showed that there was not a statistically significant difference in the amount of words students' read correctly between schools, $\chi^2(6) = 10.507, p = 0.105$ (see Table 7). Mean rank scores for each school are presented in Table 8.

Table 7
Grade 1 WRC: Results from the Kruskal-Wallis Test

Kruskal Wallis Test	
Spring WRC	
Chi-Square	10.507
df	6
Asymp. Sig.	.105
* p < .05	

Table 8
Grade 1 WRC: Mean Ranks for Each School Based on Mean Scores

	School	N	Mean Rank
WRC	1	60	343.08
	2	141	372.25
	4	58	393.92
	5	83	316.74
	6	190	328.23
	7	113	371.23
	8	54	335.71
	Total	699	

Rate of improvement. The sample consisted of 698 first grade students. The test for normality, examining standardized skewness and kurtosis, indicated the data were statistically normal. Additionally, homogeneity of variance was not significant, Levene's $F(6, 691) = 1.934, p = .073$, indicating this assumption underlying the application of ANOVA was met. An analysis of variance showed that the effect of school on rate of improvement from the fall benchmark to the spring benchmark was significant, $F(6, 691) = 6.673, p = .000$ (see Table 9).

Table 9
Grade 1 ROI: One-Way Analysis of Variance of Students' ROI by School

ANOVA						
Winter To Spring ROI						
	Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Between Groups	32.216	6	5.369	6.673	.000***	.055
Within Groups	556.045	691	.805			
Total	588.261	697				

* $p < .05$. ** $p < .01$. *** $p < .001$

Tukey post hoc tests revealed that rate of improvement was statistically significantly lower in school five than in schools one ($p = .000$), four ($p = .000$), seven ($p = .000$), and eight ($p = .006$) (see Table 10). Additionally, tukey post hoc tests demonstrated that rate of improvement was significantly lower in school six relative to school four ($p = .007$). No other school differences were statistically significant. The $\eta^2 = .055$ indicated a small to medium effect size, and further, that 5.5% of the variation in first grade rate of improvement is attributable to differences in the ROI for first grade students between the seven schools.

Table 10
Grade 1 ROI: Summary of Significant Multiple Comparisons between Schools (Tukey HSD)

Schools	Mean difference	p -value
School 4 v. School 6	0.48211	0.007**
School 4 v. School 5	0.76713	.000***
School 1 v. School 5	0.66826	.000***
School 7 v. School 5	0.55516	.000***
School 8 v. School 5	0.56494	0.006**

* $p < .05$. ** $p < .01$. *** $p < .001$

Grade Two.

Grade two data were available for six schools (i.e. school 1, 4, 5, 6, 7, & 8).

Spring words read correct. The sample consisted of 590 second grade students. The test for normality, examining standardized skewness and kurtosis,

indicated the data were statistically normal. Additionally, homogeneity of variance was not significant, Levene's $F(5, 584) = .332, p = .894$, indicating this assumption underlying the application of ANOVA was met. An analysis of variance showed that the effect of school on words read correctly at the spring benchmark was significant, $F(5, 584) = 2.42, p = .035$ (see Table 11).

Table 11

Grade 2 WRC: One-Way Analysis of Variance of Students' WRC by School

ANOVA

Spring WRC

	Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Between Groups	15654.702	5	3130.940	2.420	.035*	.02
Within Groups	755700.322	584	1294.007			
Total	771355.024	589				

* $p < .05$. ** $p < .01$. *** $p < .001$

Tukey post hoc tests revealed that the mean WRC was significantly lower for school 5 ($M = 96.07, SD = .33.891$) than for school 6 ($M = 110.36, SD = .35.819$). No other school differences were statistically significant (see Table 12). The $\eta^2 = .02$ indicated a small effect size, and further, that 2% of the variation in second grade words read correct is attributable to differences in WRC for second grade students between the six schools.

Table 12

Grade 2 WRC: Summary of Significant Multiple Comparisons between Schools (Tukey HSD)

Schools	Mean difference	p-value
School 6 v. School 5	14.288	0.03*

* $p < .05$. ** $p < .01$. *** $p < .001$

Rate of improvement. The sample consisted of 585 second grade students. The test for normality, examining standardized skewness and kurtosis, indicated the data were statistically normal. Additionally, homogeneity of variance was not significant,

Levene's $F(5, 579) = 1.645, p = .146$, indicating this assumption underlying the application of ANOVA was met. An analysis of variance showed that the effect of school on rate of improvement was significant, $F(5, 579) = 7.057, p = .000$ (see Table 13).

Table 13
Grade 2 ROI: One-Way Analysis of Variance of Students' ROI by School

ANOVA						
Fall To Spring ROI						
	Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Between Groups	6.869	5	1.374	7.057	.000***	.06
Within Groups	112.704	579	.195			
Total	119.573	584				

* $p < .05$. ** $p < .01$. *** $p < .001$

Tukey post hoc tests revealed that the mean words read correct in spring was significantly lower for school 5 ($M = 1.1811, SD = .42907$) than for school 1 ($M = 1.5725, SD = .41975$), school 6 ($M = 1.4376, SD = .47475$), and school 7 ($M = 1.3689, SD = .41791$). Additionally, Tukey post hoc tests indicated that mean rate of improvement was statistically significantly higher for school 1 ($M = 1.5725, SD = .41965$) than for school 4 ($M = 1.3012, SD = .48628$), school 7 ($M = 1.3689, SD = .41791$), and school 8 ($M = 1.3387, SD = .36385$). The $\eta^2 = .06$ indicated a small to moderate effect size, and further, that 6% of the variation in second grade rate of improvement is attributable to differences in the ROI for second grade students between the six schools (see Table 14).

Table 14

Grade 2 ROI: Summary of Significant Multiple Comparisons between Schools (Tukey HSD)

Schools	Mean difference	p- value
School 1 v. School 4	0.27131	0.011*
School 1 v. School 5	0.39135	.000***
School 1 v. School 7	0.20355	0.028*
School 1 v. School 8	0.23375	0.031*
School 6 v. School 5	0.25645	.000***
School 7 v. School 5	0.18781	0.038*

* p < .05. ** p < .01. *** p < .001

Grade Three.

Grade three data were available for seven schools (i.e. school 1, 3, 4, 5, 6, 7, & 8).

Spring words read correct. The sample consisted of 762 third grade students.

The test for normality, examining standardized skewness and kurtosis, indicated the data were statistically normal. Additionally, homogeneity of variance was not significant, Levene's F (6, 755) = .910, p = .487, indicating this assumption underlying the application of ANOVA was met. An analysis of variance showed that the effect of school on words read correctly at the spring benchmark was significant, F (6, 775) = 2.512, p = .021 (see Table 15).

Table 15

Grade 3 WRC: One-Way Analysis of Variance of Students' WRC by School

ANOVA

Spring WRC

	Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Between Groups	25571.066	6	4261.844	2.512	.021*	.02
Within Groups	1280898.283	755	1696.554			
Total	1306469.349	761				

* p < .05. ** p < .01. *** p < .001

A Tukey post hoc test revealed that mean words read correct was statistically lower for school 5 ($M = 120.33$, $SD = 44.668$) than for school 7 ($M = 139.52$, $SD = 40.746$).

The $\eta^2 = .02$ indicated a small effect size, and further, that 2% of the variation in first grade rate of improvement is attributable to differences in WRC for third grade students between the seven schools (see Table 16).

Table 16

Grade 3 WRC: Summary of Significant Multiple Comparisons between Schools (Tukey HSD)

Schools	Meandifference	p- value
School 7 v. School 5	19.184	0.031*

* $p < .05$. ** $p < .01$. *** $p < .001$

Rate of improvement. The sample consisted of 751 third grade students. The test for normality, examining standardized skewness and kurtosis, indicated the data were statistically normal. Additionally, homogeneity of variance was not significant, Levene's $F(6, 751) = 1.863$, $p = .085$, indicating this assumption underlying the application of ANOVA was met. An analysis of variance showed that the effect of school on rate of improvement from the fall benchmark to the spring benchmark was significant, $F(6, 751) = 5.169$, $p = .000$ (see Table 17).

Table 17

Grade 3 ROI: One-Way Analysis of Variance of Students' ROI by School

ANOVA

Fall To Spring ROI

	Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Between Groups	9.862	6	1.644	5.169	.000***	.039
Within Groups	238.813	751	.318			
Total	248.676	757				

* $p < .05$. ** $p < .01$. *** $p < .001$

Tukey post hoc tests revealed that the mean rate of improvement was significantly lower in school five ($M = 1.1357$, $SD = .5469$) than for school 4 ($M = 1.5117$, $SD = .62316$), for school 7 ($M = 1.441$, $SD = .56495$), and for school 8 ($M = 1.4599$, $SD = .70227$). Additionally, Tukey post hoc tests demonstrated that the mean rate of

improvement was statistically significantly lower for school 6 ($M= 1.2278$, $SD= .50286$) than in school 4 ($M= 1.5117$, $SD= .62316$), for school 7 ($M= 1.441$, $SD= .5695$), and for school 8 ($M= 1.4599$, $SD= .70227$). The $\eta^2 = .039$ indicated a small effect size, and further, that 3.9% of the variation in first grade rate of improvement is attributable to differences in the ROI for third grade students between the seven schools (see Table 18).

Table 18
Grade 3 ROI: Summary of Significant Multiple Comparisons between Schools (Tukey HSD)

Schools	Mean _{difference}	<i>p</i> - value
School 4 v. School 5	0.376	0.003**
School 4 v. School 6	0.28386	0.012**
School 7 v. School 5	0.30532	0.006**
School 7 v. School 6	0.32417	0.01*
School 8 v. School 5	0.21319	0.018*
School 8 v. School 6	0.23203	0.038*

* $p < .05$. ** $p < .01$. *** $p < .001$

Grade Four.

Grade four data were available for six schools (i.e. school 1, 3, 4, 5, 7, & 8).

Spring words read correct. The sample consisted of 533 fourth grade students.

The test for normality, examining standardized skewness and kurtosis, indicated the data were statistically normal. Additionally, homogeneity of variance was not significant, Levene's $F(5, 547) = 1.279$, $p = .271$, indicating this assumption underlying the application of ANOVA was met. An analysis of variance showed that the effect of school on words read correctly at the spring benchmark was significant, $F(5, 547) = 3.445$, $p = .004$ (see Table 19).

Table 19

*Grade 4 WRC: One-Way Analysis of Variance of Students' WRC by School***ANOVA**

Spring WRC

	Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Between Groups	33105.670	5	6621.134	3.445	.004**	.03
Within Groups	1051237.961	547	1921.824			
Total	1084343.631	552				

* $p < .05$. ** $p < .01$. *** $p < .001$

Tukey post hoc tests revealed that mean words read correct was statistically lower for school 5 ($M = 132.13$, $SD = 43.514$) than for school 1 ($M = 154.27$, $SD = 48.666$) and for school 3 ($M = 153.28$, $SD = 42.408$). The $\eta^2 = .03$ indicated a small effect size, and further, that 3% of the variation in first grade rate of improvement is attributable to differences in WRC for fourth grade students between the six schools (see Table 20).

Table 20

Grade 4 WRC: Summary of Significant Multiple Comparisons between Schools (Tukey HSD)

Schools	Mean difference	p -value
School 1 v. School 5	22.137	0.042*
School 3 v. School 5	21.153	0.009**

* $p < .05$. ** $p < .01$. *** $p < .001$

Rate of improvement. The sample consisted of 547 fourth grade students. The test for normality, examining standardized skewness and kurtosis, indicated the data were statistically normal. However, homogeneity of variance was significant, Levene's $F(5, 547) = .910$, $p = .004$, indicating this assumption underlying the application of ANOVA was not met. Additionally, further examination, including the Shapiro-Wilk test and examination of Q-Q plots tests, revealed that the data did not meet the assumption of normality. As a result, a Kruskal-Wallis test was conducted. Results showed that there was a significant difference in students' ROI from fall to

spring, $\chi^2(5) = 20.871$, $p = 0.01$ (see Table 21). Mean rank scores for each school are presented in Table 22.

Table 21
Grade 4 ROI: Results from the Kruskal-Wallis Test

Fall To Spring ROI	
Chi-Square	20.871
df	5
Asymp. Sig.	.001**

* $p < .05$. ** $p < .01$. *** $p < .001$

Table 22
Grade 4 ROI: Mean Ranks for Each School Based on Mean Scores

	School	N	Mean Rank
ROI	1	64	343.52
	3	169	281.26
	4	57	257.58
	5	70	254.54
	7	122	239.05
	8	65	287.65
	Total	547	

Grade Five.

Grade five data were available for four schools (i.e. school 4, 5, 7, & 8).

Spring words read correct. The sample consisted of 294 fifth grade students.

The test for normality, examining standardized skewness and kurtosis, indicated the data were statistically normal. Additionally, homogeneity of variance was not significant, Levene's $F(3, 290) = .438$, $p = .726$, indicating this assumption underlying the application of ANOVA was met. An analysis of variance showed that the effect of school on words read correct was significant, $F(3, 290) = 5.341$, $p = .001$ (see Table 23).

Table 23

*Grade 5 WRC: One-Way Analysis of Variance of Students' WRC by School***ANOVA**

Spring WRC

	Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Between Groups	29420.768	3	9806.923	5.341	.001**	.05
Within Groups	532524.457	290	1836.291			
Total	561945.225	293				

* $p < .05$. ** $p < .01$. *** $p < .001$

A Tukey post hoc test revealed that mean words read correct was significantly lower for school 5 ($M = 143.13$, $SD = 40.86$) than for school 7 ($M = 160.40$, $SD = 44.655$) and for school 8 ($M = 170.49$, $SD = 39.864$). The $\eta^2 = .05$ indicated a small effect size, and further, that 5% of the variation in first grade rate of improvement is attributable to differences in WRC for fifth grade students between the four schools (see Table 24).

Table 24

Grade 5 WRC: Summary of Significant Multiple Comparisons between Schools (Tukey HSD)

Schools	Mean difference	p -value
School 7 v. School 5	17.263	0.035**
School 8 v. School 5	27.358	0.001**

* $p < .05$. ** $p < .01$. *** $p < .001$

Rate of improvement. The sample consisted of 290 fifth grade students. The test for normality, examining standardized skewness and kurtosis, indicated the data were statistically normal. Additionally, homogeneity of variance was not significant, Levene's $F(3, 287) = 1.832$, $p = .142$, indicating this assumption underlying the application of ANOVA was met. An analysis of variance showed that the effect of school on rate of improvement was significant, $F(3, 287) = 4.757$, $p = .003$ (see Table 25).

Table 25
Grade 5 ROI: One-Way Analysis of Variance of Students' ROI by School

ANOVA

FallToSpringROI

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.188	3	1.063	4.757	.003**
Within Groups	64.118	287	.223		
Total	67.306	290			

* $p < .05$. ** $p < .01$. *** $p < .001$

Tukey post hoc tests revealed that the mean rate of improvement was significantly lower in school five ($M = .8124$, $SD = .45146$) than for school 7 ($M = 1.0663$, $SD = .52163$) and for school 8 ($M = 1.0357$, $SD = .38266$). The $\eta^2 = .048$ indicated a small effect size, and further, that 4.8% of the variation in first grade rate of improvement is attributable to differences in the ROI for fifth grade students between the four schools (see Table 26).

Table 26
Grade 5 ROI: Summary of Significant Multiple Comparisons between Schools (Tukey HSD)

Schools	Mean difference	p -value
School 7 v. School 5	0.25388	0.002**
School 8 v. School 5	0.22334	0.033*

* $p < .05$. ** $p < .01$. *** $p < .001$

Summary. The second research question sought to examine the extent to which the fidelity of RtI implementation relates to differences in students' ability to read accurately and fluently for one minute, measured by words read correct (WRC) and rate of improvement (ROI), between elementary schools, in grades one through five. The research question was analyzed in two steps. First, some general conclusions can be made about the differences in student outcomes between schools. A total of 25 significant school differences in student reading outcomes emerged across the five grade levels. Greater differences in student reading outcomes between schools were

noted on the rate of improvement (ROI) variable than on the words read correctly (WRC) variable. Specifically, 19 significant differences between schools were found on the ROI variable and six significant differences were found on the WRC variable. On the WRC variable, school 5 was consistently the lower performing school across grades. Schools 1, 3, 6, 7, and 8 had greater student reading outcomes than school 5 on the WRC variable.

The results were more varied between grades on the ROI variable. In grade one, schools 1, 4, 7, and 8 were higher performing than school 5. In grade two, school 1 was a higher performing school than schools 4, 5, 7, and 8. In grade three, schools 4, 7, and 8 were higher performing than schools 5 and 6. In grade four and grade five, school 5 was consistently the lower performing school, compared to schools 1, 3, 7, and 8. Finally, some school differences were recurrent across grade levels. Specifically, school 7 performed better than school 5 at four grade levels (i.e., grades 1, 2, 3, 5), school 1 performed better than school 5 at three grade levels (i.e., grades 1, 2, 4), and school 8 performed better than school 5 at three grade levels (i.e., grades 1, 3, 5).

Second, *qualitative, exploratory* inferences were made to investigate the relationship between RtI implementation and differences in student outcomes. Table 27 was created to compare the significant differences in student outcomes related to ORF with schools fidelity of RtI implementation. The left column of the table lists the significant differences in student reading outcomes by grade level and the first school listed in the pair performed higher on student outcomes. The remaining columns represent the five essential components of RtI and the total RtI score. The checkmark

represents that the first school listed in the pair had a higher fidelity of implementation score than the second school listed in the pair. When reading across each row, the reader can determine on which components the school that performed better on reading outcomes related to ORF also had higher RtI fidelity scores.

The following are qualitative inferences were made based on the results listed in Table 27. The three schools with the highest total RtI scores (i.e. schools 1, 2, & 3) did not have significantly lower student reading outcomes than any other school, at any grade level. However, it is notable that the schools with the greatest RtI total scores also had the least amount of grade levels available for the analysis. All schools that performed significantly better than another school had at least one component, and/or RtI total, that was greater than the lower scoring school. Next, an analysis of all 25 significant differences revealed that 92% of the schools scoring higher on student reading outcomes had higher infrastructure and support ratings, 80% of the schools scoring higher on student reading outcomes had higher RtI total ratings, and 72% of the higher school had higher fidelity and evaluation ratings. In other words, one possible explanation for the trend observed in the data is a positive association between significantly better student reading outcomes related to oral reading fluency (ORF) and higher ratings on the infrastructure and support component, the fidelity and evaluation component, and the RtI total score. Similarly, 72% of the schools that scored significantly higher on student reading outcomes had three or more components, including RtI total, that were rated higher than the lower performing school, with half (i.e., 50%) of that 72% having higher ratings for all 5 components and for RtI total (i.e., all 6 aspects of RtI). In addition, higher ratings on MTSS appear

to be least related to differences in student outcomes. Finally, it is possible that assessment and data based decision making (DBDM) also impacted differences in student outcomes between schools, yet the current study was unable to address this because of the similarity of implementation between schools. It is important to note that these inferences are speculative and exploratory in nature.

Although school 5 and school 8 have similarly low RtI ratings, school 8 had significantly better student outcomes than did school 5 at three grade levels (i.e., grade 1, 3, 5). This finding suggests that other variables may be impacting student outcomes. It is possible that demographic variables, such as the percentage of students receiving free/reduced lunch, the percentage of nonwhite students, and/or the percentage of students receiving special education services, influenced the difference in student outcomes, as school 5 is the most diverse school on all of these factors. In addition, school 5 and school 4 have the greatest overall diversity, with similar percentages of students receiving free/reduced lunch and the percentage of nonwhite students. Although these two schools have a greater percentage of nonwhite students, school 4 has approximately half the nonwhite students as school 5. Despite these commonalities, the results from the ANOVA's were quite different for these two schools, as school 5 consistently had lower student outcomes, and school 4 had significantly higher student outcomes than school 5 in grades 1 and 3. Finally, school 4 received higher ratings than school 5 on all five RtI components, and on RtI total. One possible explanation for the significant differences in student reading outcomes related to ORF between school 4 and school 5 is the greater percentage of nonwhite students in school 5. Another possible explanation for the significant difference in

student outcomes between schools may be that implementation of RtI based educational service delivery may have an impact on student outcomes above and beyond demographic characteristics.

Table 27
Comparison Of Significant Post-hoc Tukey Tests and Fidelity of RtI Implementation

	Assessment	DBDM	MTSS	Infrastructure & Support	Fidelity & Evaluation	RtI Total
Grade 1						
4 v. 6 WRC	✓	✓	✓	✓	✓	✓
4 v. 5 ROI	✓	✓	✓	✓	✓	✓
1 v. 5 ROI	✓	✓	✓	✓	✓	✓
7 v. 5 ROI		✓		✓	✓	✓
8 v. 5 ROI				✓		
Grade 2						
6 v. 5 WRC					✓	✓
1 v. 4 ROI			✓	✓		✓
1 v. 5 ROI	✓	✓	✓	✓	✓	✓
1 v. 7 ROI	✓		✓	✓	✓	✓
1 v. 8 ROI	✓	✓	✓	✓	✓	✓
6 v. 5 ROI					✓	✓
7 v. 5 ROI		✓		✓	✓	✓
Grade 3						
7 v. 5 WRC		✓		✓	✓	✓
4 v. 5 ROI	✓	✓	✓	✓	✓	✓
4 v. 6 ROI	✓	✓	✓	✓	✓	✓
7 v. 5 ROI		✓		✓	✓	✓
8 v. 5 ROI				✓		
7 v. 6 ROI	✓	✓		✓		✓
8 v. 6 ROI	✓			✓		
Grade 4						
1 v. 5 WRC	✓	✓	✓	✓	✓	✓
3 v. 5 WRC	✓	✓	✓	✓	✓	✓
Grade 5						
7 v. 5 WRC		✓		✓	✓	✓
8 v. 5 WRC				✓		
7 v. 5 ROI		✓		✓	✓	✓
8 v. 5 ROI				✓		
Total ✓	48%	64%	44%	92%	72%	80%

✓ = the first school listed in the pair received a higher score

Discussion

The primary purpose of the current study was to explore several aspects of the fidelity of implementation of RtI, including (1) the extent to which elementary schools were implementing RtI as measured by the RtI fidelity of implementation worksheet and rubric, (2) how differences in implementation could be characterized, and (3) to what extent RtI implementation was related to differences in student-level outcomes in grades one through five. These topics are important because various RtI models exist, and are being implemented and used for high-stakes decision making about students on a nationwide scale. Although numerous studies have demonstrated a relationship between individual components of RtI and positive student outcomes, few studies have examined the link between student outcomes and implementation of RtI as a coherent system.

In the current study, the extent of RtI implementation, defined by five essential components, was evaluated in eight elementary schools from the same collaborative education district. The five components of RtI subjected to fidelity examination were: assessment, data-based decision making (DBDM), multi-tiered system of supports (MTSS), infrastructure and support, and fidelity and evaluation. In addition, summing all of the points earned from ratings of each of the five essential components derived a “total RtI score”. Each of the eight elementary schools scored in the *high implementation range* for assessment and DBDM, thus indicating similarly high quality of implementation across schools. Three elementary schools scored in the *high implementation range* for MTSS, fidelity and evaluation, and total RtI, while five elementary schools scored in the *low implementation range*, indicating differences in

quality of implementation across schools. Only one school scored in the *high implementation range* for infrastructure and support. Further, the greatest variability in ratings was observed on the *fidelity and evaluation* component, followed closely by the *infrastructure and support* component. Clearly, there are similarities and differences in the quality of RtI implementation across schools despite the schools having similar professional development and supports around the topics and practices of RtI.

Examining the relationship between fidelity of RtI implementation and student reading outcomes related to ORF was exploratory in nature and qualitative inferences were used to speculate about the data. Regarding the relationship between RtI implementation and student outcomes, the data indicated mixed results as follows. Some of the results demonstrated a trend indicating there was a possible positive association between RtI implementation and student reading performance and growth. For example, most of the data support a relationship between significantly higher student reading outcomes and higher ratings on the *infrastructure and support* component, the *fidelity and evaluation* component, and the total RtI score. In addition, the majority of the data (i.e., 72%) support a relationship between high quality implementation on three or more components and higher student outcomes. Finally, schools that scored in the *high implementation range* on total RtI had student outcomes that were equal to or better than student outcomes from any other school across grade levels. Despite indications of a positive association between fidelity of RtI implementation and student outcomes, other results failed to support a clear relationship between the two variables. For example, the schools with the highest

quality of RtI implementation were not significantly different, statistically, on student reading outcomes as compared with schools with lower quality RtI implementation. This was surprising, and counter to the researcher's expectations. It is possible that this was related to these schools providing the least amount of data. For example, school 2 only provided data for grade one. Another possible explanation for this finding is that student reading outcomes related to ORF were generally higher at the beginning of the school year for these schools, and ceiling effects of the measure may have limited the rate of improvement measure. Second, the two schools with the lowest RtI implementation ratings overall were found to have significantly different student reading outcomes at multiple grade levels. That is, school 8 performed better than school 5 in grades one, three, and five. In summary, some of the data support a positive relationship between RtI implementation and student reading outcomes, whereas, some of the data support the conclusion that the relationship is inconsistent.

A greater understanding of fidelity of RtI implementation, measured by the quality of RtI implementation, and its' relationship to student reading outcomes related to ORF may help schools understand which components of RtI are critical to changes in student outcomes. Additionally, it may guide schools toward the optimal use of time and resources to positively impact student outcomes through RtI. The current study is one of the first studies to examine the fidelity of RtI implementation as a system of coordinated components.

How Are These Results Similar To And Different From Those Of Previous Studies?

The first research question described RtI implementation in eight elementary schools. In the current study, assessment and data based decision making (DBDM)

were found to have the highest quality of implementation, which was consistent with prior research (Sharp et al., 2016). Bailey (2014) also found that assessment was the area of RtI with the highest quality of implementation; however, DBDM was found to be the most challenging for schools and was implemented with low quality. In the current study, the variables *infrastructure and support* and *fidelity and evaluation* were the two areas of RtI implementation that received the lowest scores. These results were consistent with Sharp et al., (2016) who found that professional development (i.e., supports) and fidelity of implementation received the lowest scores. Although there were similarities between the current study and Sharp et al., (2016) regarding the areas of RtI with the highest and lowest quality of implementation, it is notable that the scores from the current study were generally higher (i.e., ranging from 3.37 to 4.86) than those of Sharp et al. (i.e., ranging from 1.95 to 2.97), indicating an overall greater quality of RtI implementation. It is possible that the discrepancy in scores is related to the number of years the schools were implementing RtI, as in Sharp et al., (2016) the mean number of years was 3.97 and in the current study the schools were implementing RtI for approximately 20 years. Similarly, the mean overall RtI implementation score in the current study was 79% compared to 55% found in a prior study, with a similar sample size, using an alternative RtI implementation integrity measure (Noltemeyer & Sansosti, 2012). Next, the current multi-tiered system of support (MTSS) appeared to be least associated with student reading outcomes related to oral reading fluency, which is inconsistent with prior research. For example, Noltemeyer and Sansosti (2012) found that academic implementation of MTSS was a significant predictor of student reading outcomes as

measured by oral reading fluency. Finally, the current study is the first to examine a “total RtI score” that incorporated all components of RtI implementation. Taken together, the variation in RtI implementation scores across studies reinforces the notion that RtI implementation varies across schools, districts, and states.

Prior studies have suggested that a relationship exists between overall fidelity of RtI implementation and student reading outcomes (Noltemeyer & Sansosti, 2012), and between data-based decision making and student reading outcomes (Sharp, Sanders, Noltemeyer, Hoffman, & Boone, 2016). As discussed above, the results from the current study have preliminary indications for mixed results regarding the relationship between RtI implementation and student reading outcomes. Mixed results may have been found due to other factors aside from the fidelity of RtI implementation such as student demographic characteristics and teachers’ experience implementing RtI. Further examination of the results also leads to some interesting exploratory *speculation* for discussion on fidelity of RtI implementation.

For example, one issue that emerged from the results is related to diversity, and it is described in the context of the relationship between schools 5 and 8 and the relationship between schools 4 and 5. First, school 5 and school 8 acquired similar ratings on the RtI implementation integrity measure, which were characterized by ratings that fell in the high implementation range for assessment and DBDM, and ratings that fell in the low implementation range for MTSS, infrastructure and support, fidelity and evaluation, and total RtI score. Further, these two schools received the lowest total RtI ratings compared to the other six schools. Considering the commonalities in the RtI ratings, it was expected that these schools would have similar

student reading outcomes. However, student reading outcomes in school 8 were significantly higher than student reading outcomes in school 5 in grades one, three, and five. Further investigation into these differences revealed that these two schools varied greatly on demographic characteristics. For example, the lower performing school 5 had 64.6% students receiving free/reduced lunch compared to 34% in school 8. Similarly, school 5 had 30% nonwhite students compared to 6.6% in school 8. Two prior studies have explicitly examined the relationship between RtI implementation and student reading outcomes while controlling for demographic variables and found mixed results. More specifically, one study found that the percentage of economically disadvantaged students and the percentage of non-minority students were not significant predictors of student outcomes; however, two schools were removed from the sample to make the sample more homogeneous (Noltemeyer & Sansosti, 2012). In contrast, another study found that the percentage of economically disadvantaged students was a significant predictor of student outcomes accounting for more variability in scores than all RtI implementation components (Sharp et al., 2016). Thus, it is important to consider whether the observed differences in reading outcomes between school 5 and 8 might be attributable to differences in students SES and diversity.

On a related note, school 4 and school 5 were the most diverse participating schools in the study. For example, school 5 had 64.6% students receiving free/reduced lunch compared to 61% at school 4. In addition, school 5 had 30% nonwhite students compared to 14.9% in school 4. The remaining six schools had between 4.9% and 6.6% nonwhite students, and between 23.9% and 47.9% students receiving

free/reduced lunch, which demonstrates the greater diversity in schools 4 and 5 compared to the other schools. School 5 and school 4, however, differ in their fidelity of RtI implementation scores. School 4 earned higher ratings on all RtI components, and total RtI score, than school 5. Both schools fell in the high implementation range for assessment and DBDM, yet school 4 also fell in the high implementation range on the *infrastructure and support* component. Results from the study revealed that student reading outcomes in school 4 were significantly higher than student reading outcomes in school 5 in grades one and three. Despite the diversity in school 4, student reading outcomes were significantly higher than school 5. The results from the current study are consistent with prior studies suggesting that demographic variables should be accounted for when examining the relationship between RtI implementation and student outcomes.

The second issue that emerged from the results is related to grade level. Some of the differences in student outcomes found between elementary schools were consistent across multiple grade levels, and some were not. School 4 had significantly higher student reading outcomes than school 5 and school 6 in grade one and in grade three. School 1 had significantly higher student reading outcomes than school 5 in grade one, grade two, and grade four. School 8 had significantly higher student reading outcomes than school 5 in grade one, grade three, and grade five. Finally, school 7 had significantly higher student reading outcomes than school 5 in grade one, grade two, grade three, and grade five. The remaining significant differences in student reading outcomes between elementary schools emerged at only one grade level. It is unclear why the results are more robust, or consistent across grade levels,

for some differences in student reading outcomes between schools and not others. These data suggest that further investigation into the relationship between fidelity of RtI implementation and student reading outcomes, by grade level, should be addressed in more detail.

Another issue related to the current study is the use of oral reading fluency (ORF) as the dependent measure. There are limitations associated with using only ORF based scores to evaluate student reading outcomes in relation to RtI. First, despite being developed as a measure of rate and accuracy (i.e. fluency), some researchers have found that teachers perceive ORF as a measure of accuracy and speed, or students' ability to read quickly in one minute (Deeney & Shim, 2016). Second, some researchers have argued that oral ready fluency measures such as R-CBM (i.e. measure used in the current study) do not measure the full construct of oral reading fluency because the measure does not directly address reading comprehension and expression (Hosp & Suchey, 2014). Finally, ORF measures do not take into account a students' ability to read for longer periods of time, which might be influenced by factors such as attention, persistence, and stamina. That said, it is also the case that there is no single measure of reading that comprehensively measures the full range of reading fluency characteristics. There is a need to develop new reading measures that are efficient and sensitive to changes in student learning at different time points during the school year. These measures should measure multiple facets of oral reading fluency including expression, accuracy, comprehension, and persistence. These are all important factors that impact students' ability to learn in school. Also in future research, other variables should be considered to include more robust reading

outcomes that could potentially capture the impact of the challenging and complex process of RtI implementation.

Overall, the results from the current study indicate that the relationship between fidelity of RtI implementation and student reading outcomes is not direct. Many studies have shown a clear and direct relationship between student reading outcomes and reading instruction (Joseph, 2014). Although many studies have claimed that RtI-based practices improve student reading outcomes, those studies have not provided implementation integrity data, limiting the ability to conclude that improved student outcomes were a result of RtI implementation (Dexter, Hughes, & Farmer, 2008). It is also the case that RtI based educational service delivery is grounded in, ideally, the provision of high quality, differentiated instruction. Thus, it appears that other factors mediate and/or moderate the relationship between fidelity of RtI implementation and student reading outcomes related to oral reading fluency (ORF), and more research is needed to understand the circumstances and nuances that mediate/moderate the relationship. For example, use of other reading outcome variables such as reading comprehension may produce different results. This conclusion is more consistent with the extant literature base, which contends that the multilevel and systematic nature of RtI, with all of its moving parts, makes it challenging to examine as a coherent system (Keller-Margulis, 2012).

Limitations

Although this study yielded useful and important information with respect to the quality of RtI implementation and its relationship to student reading outcomes, there are limitations as well. Numerous limitations resulted from the use of an extant

data set. First, a primary limitation was the small sample of schools ($n = 8$) available for analysis. With such a small sample size, there was limited variation in the quality of RtI implementation and, consequently, meaningful inferential analyses with RtI as the dependent variable could not be conducted. This means that differences in RtI implementation between schools could not be analyzed statistically. In addition, the schools with the highest levels of RtI implementation across the five components had the least amount of student reading outcome data available. As a result, a direct comparison of a high quality implementation group and a low quality implementation group was not achievable. Further, the lack of variation in the quality of RtI implementation impacted the reliability analysis for the RtI measure (see Appendix D).

A second limitation of the extant data set was the multiple (or inconsistent) levels of available information. More specifically, three levels of data were provided: school level RtI implementation information, individual student outcome information, and aggregated student demographic information. With the data only available at these different levels, the analyses that could be conducted were restricted. One limitation to the analysis was the inability to include demographic variables in the analysis because in order to enter the variables as covariates, the demographic information must be available at the same level as the dependent variable (i.e., student level). However, the dependent variables (i.e., ROI and WRC) were individual student level information and the demographic variables were aggregated student level information. Accounting for demographic variables would have improved the internal validity of the study by increasing the confidence in attributing differences in student outcomes to RtI

implementation (Cook & Campbell, 1976). Although the student outcome data could have been aggregated (i.e., by calculating the mean) to match the level of the demographic information, the school sample size was too small to yield useable information, as there would have been only eight data points.

Another limitation that resulted from the different levels of information was the inability to conduct a regression analysis. This is because regression analysis requires the level of the data to be consistent across the independent variable/s and the dependent variable/s (Harlow, 2014). However, the extant data set consisted of school level RtI implementation as the independent variable/s and individual student level data as the dependent variable/s. Similar to the demographic information, if the individual student level data was aggregated (i.e., the mean) to school level, the sample size would be too small to run the regression analysis. In multiple regression, the guideline for sample size has been suggested to be approximately 30 participants (i.e., data points) per predictor variable (VanVoorhis & Morgan, 2007). The current study proposed to examine six predictor variables, which would require a much larger sample size.

A third limitation of the extant data set was the availability of oral reading fluency (ORF) as the sole student reading outcomes variable. The current study may have been more meaningful if multiple student reading outcome variables, or other achievement indicators were able to be incorporated. For example, a reading comprehension measure such as AIMSweb MAZE, developmental reading assessment (DRA), or state achievement test data may have added more information. Unfortunately, no one measure is available to measure the full construct of reading

fluency (Hosp & Suchey, 2014). Consequently, it may be beneficial to replicate the current study with the use of multiple dependent variables that reflect student reading outcomes.

Finally, the extant data set does not include a control for comparison. In the current study, a control could have manifested in two ways. First, schools with limited or no experience implementing RtI may have been used as a control group. Second, student reading outcome data prior to RtI implementation may have been used as a control. Examining prior student outcome data would have provided an estimate of student growth in the same schools before RtI was introduced and allowed for comparison of student growth before and during RtI implementation. As such, including some type of comparison would have increased the overall internal validity of the study because a direct comparison, such as the two mentioned above, would increase confidence that differences in student reading outcomes were a result of the RtI implementation rather than some alternative variable/s (Cook & Campbell, 1979).

Implications for Practice

RtI as a complete and coherent system. RtI is a comprehensive framework for providing multiple levels of service delivery to meet individual student needs. Assessment, DBDM, MTSS, infrastructure and supports, and fidelity and evaluation have been identified as the five essential components that are critical to defining an RtI model (NCRTI, 2014). Consequently, all five components must be formally in place to attribute changes in student outcomes to implementation of RtI, and to apply RtI to special education eligibility decisions. However, there is no single method to implement RtI appropriately, and various models of RtI exist to fit the structure and

culture of school districts and individual schools; some adaptation is permitted and inevitable when implementing programs or interventions in applied settings. The extant literature on RtI has focused on the implementation of the individual components of RtI; however, when implemented as designed, the crux of RtI is the relationship between all components being implemented well. For example, the information collected from assessments is used to advise decisions about student learning (DBDM), and decisions formulated from DBDM leads to movement between multiple tiers of student supports (MTSS). Further, *infrastructure and supports* either assist and encourage, or create obstacles for implementing assessment, DBDM, and MTSS.

Arguably, the most important aspect of the RtI framework is *fidelity and evaluation* (i.e. treatment integrity). *Fidelity and evaluation* is the underlying piece that unifies the components. As RtI is a continuous cycle of information gathering, decision making, and evaluation, the effectiveness of the framework is reliant on the integrity of each step. Specifically, if the information gathered is not reliable and valid, then decisions about student learning are likely to be inaccurate, and as a result, the tiers of support or movement between tiers may not be appropriate. In other words, if the implementation fidelity of the process breaks down in any one component or aspect within RtI, then the entire system is likely to be flawed (VanDerHyden et al., 2007). In this regard, the current study found that a trend emerged supporting a positive relationship between *fidelity and evaluation* scores and student reading outcomes. Given the importance of fidelity of implementation to the effectiveness of the overall RtI framework, efforts should be made to support the integrity of

implementation. Limited research is available to describe the best methods for supporting the implementation of school-wide programs such as RtI, so it is advantageous to think about ways to generalize what has been learned from the intervention plan implementation research and think about how it can be applied on a broader scale.

What do we know about intervention plan implementation in schools?

Implementing educational programs and interventions in schools is more complex than in a controlled research setting. Previous research has revealed that fidelity of implementation is reduced in applied settings, such as schools, and implementation of interventions is often poor and declines over time, resulting in varied levels of implementation (Noell & Gansle, 2016; Odom et al, 2010). Numerous conceptualizations for the measurement of fidelity of implementation of interventions or programs have been theorized and examined in applied settings (Noell & Gansle, 2016). Despite the rationale supporting more in depth methods for measuring fidelity of implementation (i.e. quality, program differentiation), one of the more basic methods has received the most support. ‘Procedural adherence’ is the recommended method for measuring fidelity of implementation of interventions in the school setting because it has been found to be the most practical (i.e. reliable and feasible) and successful (i.e. related to changes in student achievement) method (Sanetti & Kratochwill, 2012). To collect procedural adherence data, an educator uses a predetermined list of discrete, observable steps of the intervention and determines whether each step occurred or did not occur during administration. As a result, the percentage of procedural steps completed is calculated and determines the level of

integrity to the intervention process. Mixed results have been found regarding the relationship between procedural adherence and student outcomes (i.e. academic and behavior) . For example, it is likely that all steps of an intervention are not equally important, and the importance of the steps may vary over the course of implementation (Schulte, Eaton, & Parker, 2009; Noell & Gansle, 2006). In another example, declines in procedural adherence have been found to reduce child compliance for some students while other students demonstrate growth (Leon et al. 2014). Leon et al., 2014 indicates that variability in response to declines in procedural adherence exists across individuals. It is possible that some students require less procedural adherence than other students to benefit from an intervention; however, the underlying reason why that occurs is unclear. Further, the types of administration errors (i.e. omitting versus incorrectly administering) may impact the effectiveness of the intervention (Leon et al. 2014).

Given the variation in individual response to the same intervention, it is important for implementation to be monitored and supported to achieve the greatest success from the intervention. Research suggests that direct methods such as observation and permanent product review are more effective than indirect methods such as self-report to monitor intervention implementation (Sanetti & Kratochwill, 2012). Regarding support for implementation, performance feedback is the most successful method for supporting fidelity of intervention implementation supported by the literature (Noell & Gansle, 2016). Other methods, including increased training and the quality of relationship between the practitioner and the student/s, have not been consistently supported in the literature (Noell & Gansle, 2016). Performance feedback

involves direct measurement of the intervention implementation and a follow-up meeting to provide feedback regarding the procedural adherence to the intervention. In one study, Coddling, Feinberg, Dunn, and Pace (2005) found that performance feedback was effective when used on a daily basis, and when the feedback was faded up to two weeks. In addition, to facilitate implementation of interventions in applied settings, resources such as time, space, personnel, and training need to be shifted to meet the needs of the intervention (Maltzman, 2016).

Based on the available information on fidelity of implementation of interventions or programs in schools, it is clear that fidelity of implementation is important, yet achieving high levels of fidelity is a complex undertaking. Rather, research suggests that fidelity of implementation is essentially human behavior and, in schools, the behaviors of educators are maintained by the environment and context within which teaching occurs (Sanetti & Kratochwill, 2012; Maltzman, 2016).

Context of RtI implementation. As a prevention and early intervention model, the RtI approach is predicated on the belief that student learning difficulties emanate from a mismatch between instructional activities/resources and student needs, and that all students can learn and make progress if provided with the appropriate supports (Burns & Gibbons, 2008). In addition, the implementation of RtI models is an evolving process that encourages flexibility in school systems and infrastructure to effectively meet the needs of students. Further, schools are busy and complex environments, and educators have an overwhelming number of responsibilities to accomplish on a day-to-day basis including the implementation of interventions. The moving parts of RtI implied in the research (i.e. evolving systems and infrastructure,

numerous responsibilities) make it more challenging to define, measure, and evaluate implementation.

The current study applied the concept of fidelity of intervention or program implementation at a broader level (i.e., RtI framework) than that of individual interventions, and incorporated the context of RtI implementation in the evaluation. Results suggested a general trend emerged supporting a positive relationship between integrity of *infrastructure and supports* and student reading outcomes, which supports the notion that systems level supports are necessary for successful RtI implementation and, more specifically, for the integrity of interventions implemented within all three tiers of RtI. In addition, the current study provides support that the fidelity of RtI implementation is context dependent and the supports that are most effective for one school may not be the most effective supports for another school. This is apparent in the varying levels of implementation within the same collaborative school district, which has the same *infrastructure and support* services available to the elementary schools. These results raise the question of whether collecting fidelity of implementation data at the broader level is feasible for typical schools and school personnel.

Feasibility and recommendations for collecting RtI fidelity of implementation data. The literature clearly demonstrates that fidelity of implementation is related to the success or failure of interventions or programs in schools (Sanetti & Kratochwill, 2010). Despite this knowledge, many questions remain unanswered regarding the most effective methods for supporting implementation. It is likely that questions remain unanswered because schools are

complex behavioral systems, in which the implementation of interventions is one out of a number of competing demands. In contrast to a controlled research setting, educators are limited by the resources, infrastructure, and support systems available to them, which contributes to the necessity of implementation adaptation and flexibility that has been found to be inevitable in schools (APA, 2005). Given the intricacies of fidelity of intervention implementation at the individual and small group level, examining fidelity of implementation at the broader, RtI, level may be intimidating and may appear impractical in both research and practice.

Fidelity of implementation is arguably the most important aspect of the RtI framework because it can be, and should be applied on a small scale as well as a large scale when RtI is implemented as designed. On a small scale, fidelity of implementation information can be collected for tier I whole class instruction, tier II group interventions, tier III individualized interventions. Such information may also be useful in the monitoring of the data based decision making (DBDM) process. Examining the components of RtI and the total RtI process at a broad level, not only allows schools to attribute changes in student outcomes to the implementation of RtI, it also informs schools with respect to how to improve educational level practices (Schulte, Easton, & Parker, 2009). For example, if steps of an intervention were omitted, the data collected demonstrates and informs educators of the omission that may not have been realized without the data. In addition, RtI teams make important decisions about hypothesized reasons for student difficulties and student response to intervention. These decisions are more valuable when fidelity of implementation data is available to review. Without this information, the cycle of information gathering,

decision making, and evaluation begins to break-down. For these reasons, it is advisable for schools to collect fidelity of implementation data.

Ideally, it is best to collect fidelity of implementation data at both the broad level (i.e. RtI framework) and the small-scale level (i.e. programs and interventions in tiers I, II, III; DBDM). However, accomplishing this task in a complex school environment is a challenge. Evaluating fidelity of implementation of individual or group interventions is more likely to be feasible for schools than to collect fidelity of implementation data of the RtI framework. Additionally, information at the small-scale level is necessary for evaluating the overall effectiveness of RtI. Therefore, if integrity data is only collected at one level, the small- scale level is likely the best choice. Without this information, the fidelity of implementation data for the RtI framework is limited and less meaningful. However, if schools only focus on one level of fidelity of implementation, then the entire system may be flawed because the components are not mutually exclusive, and school personnel may remain unaware that particular aspects (i.e. professional development, evaluation, use of evidence-based practices etc.) within the components are missing, and more important, what directions are indicated for improving practice. The current study suggests that the two components of RtI most likely to impact student outcomes are 1) *infrastructure and supports* and 2) *fidelity and evaluation*. As a result, it is important for research to examine more feasible ways for schools to collect fidelity of implementation data, and to support intervention implementation, as well as, RtI implementation.

The semi-structured interview and the rubric used in the current study measured the quality of RtI implementation. These measures extended beyond

procedural adherence to include descriptions for multiple levels of implementation (i.e. 0, 1, 2, 3, 4, 5) and required a discussion with key RtI school personnel to collect data. The information collected in the structured interview was used to complete the rubric and it is likely that this process is not feasible for schools. However, the rubric could be modified to be more practical as recommended in the literature. First, the descriptions for a rating of '5' on all items could be modified into a procedural adherence checklist. Second, performance feedback could be used to support implementation if the data collected using the checklist is reviewed with RtI teams, leadership teams and/or problems solving teams. Finally, it would be most feasible to collect fidelity of implementation data at both the small-scale level and the broad level if schools are adapting systems and shifting resources to support RtI implementation, as the literature advocates is necessary for successful and effective RtI implementation.

What teachers need to feel supported in implementing RtI. Based on the results from the current study, it is clear that teachers need support to implement RtI with fidelity. It is important that support mechanisms set a tone of encouragement and assistance, rather than one of evaluation and criticism. Teachers need strong leadership from administration and RtI teams (i.e. leadership teams, problem solving teams), as well as, optimism and acceptability of the RtI process. Implementing RtI fully and well, involves a process that occurs over multiple years, and this understanding is crucial in helping school personnel engage in high quality implementation one step at a time. Teachers also need access to appropriate resources to assist with implementation. For example, it would be helpful for teachers to have a list of

evidence-based practices for tier II and tier III interventions available for their use. Teachers should not be expected to develop materials to carry out interventions in their classroom. It would also be helpful to have data specialists and intervention specialists to assist with the implementation of tier II and tier III interventions. Unfortunately, it is too commonly seen in schools that general education teachers are expected to carry out these small group and individualized interventions while also managing their classrooms and providing appropriate instruction for large numbers of children. Support mechanisms should allow teachers to feel comfortable and confident with their role in the RtI implementation process, and not add to their numerous responsibilities in a classroom. Similarly, professional development opportunities should be lead by the teachers and the school personnel who are actively engaged in the RtI implementation process. Professional development should be directly applicable to teachers' practices/roles and provide for hands-on opportunities to practice newly learned skills. Finally, making mistakes and learning from mistakes should be encouraged in schools. Too often observations of teachers are associated with evaluation and accountability, rather than an opportunity to learn and improve. Schools would benefit from creating a culture in which teachers feel comfortable trying new practices or techniques and one in which receiving feedback about their practices and implementation is viewed positively. Although these are some examples of how to support teachers, this is not an exhaustive list. It is important for researchers to learn more about the supports teachers are interested in and to examine the most effective ways to implement those supports in schools.

Future Directions

In the current study, RtI implementation in eight elementary schools was described and differences in student reading outcomes related to oral reading fluency (ORF) were discussed within the context of RtI implementation. Mixed results were found, as some of the data supported a trend indicating that greater RtI scores were related to greater student outcomes and some of the data suggested that greater student outcomes were not related to RtI scores. However, the sample size was small and the extant data set did not include the levels of information necessary to run an anticipated prediction model. As a result, the current study was not able to directly examine the relationship between the implementation of RtI components, and total RtI, and student reading outcomes. These relationships could be studied in future research with some adjustments to the data set. First, individual demographic information would allow the researcher to control for differences in student outcomes due to demographics. The current study had access to school-level demographic information; however, valuable information is lost when aggregating information that has substantial research supporting its' relationship to student outcomes. Further, having this information available would allow researchers to disaggregate findings for subgroups and determine if the impact of RtI implementation varies by demographic characteristics.

Second, with a larger sample size, or a greater number of schools in the data set, multi-level modeling is another option for analyzing the data with the purpose of prediction. Multi-level modeling would additionally begin to inform if some components are more important than other components within the RtI framework, or if the importance of components are related to other factors such as demographics

characteristics of the school. Third, to directly assess whether high implementation of RtI is more effective than low implementation of RtI, the data set should include greater variability across schools in total RtI implementation, as well as variability (i.e. numerous high implementation and numerous low implementation) between schools within each essential component. In the current study, all schools fell in the high implementation range on assessment and DBDM, limiting the variability on these components. Examining this topic with two, or three, levels of implementation and being able to directly compare them would be valuable to the literature and to the overall discussion of RtI effectiveness. Studying the varying levels within components as well will provide valuable information regarding whether some components are more important than others and if weighting the components is a better representation of the framework rather than treating all components equally when examining RtI as a system of coordinated components. Additionally, future research should look more carefully at the *fidelity and evaluation* component, and consider measuring this component in greater detail to account for 1) the overall importance of integrity within the RtI framework and 2) demonstrating that the components of RtI are interrelated through the concept of integrity to the process.

Third, future studies should consider what other factors may influence fidelity of RtI implementation and control for these factors in the analysis. It is important to consider what other factors influence RtI implementation in order to gain a better understanding of the relationship between fidelity of RtI implementation and student reading outcomes. Other factors may include teacher demographic characteristics such as the number of years with experience implementing RtI, level of education, and

beliefs about the RtI process. Similarly, since RtI implementation is a multi-year process, it would be helpful to consider any changes in staff during the implementation years. Additionally, given what is known about the impact of student demographic characteristics on overall student reading outcomes, it will be important to evaluate/monitor to any changes in student demographic characteristics while RtI implementation is taking place over time.

Finally, future studies examining the relationship between RtI implementation and student outcomes should also consider either: (a) only including students who are low performing in terms of educational achievement, or (b) include separate analyses for students performing at lower and higher levels of achievement. In other words, although RtI is a prevention model that was developed to benefit all students in a school, it is may be that RtI is most beneficial for students who are in need of supplemental instructional supports. If this is indeed the case, then potential ceiling effects can be avoided and/or better understood.

Appendix A: Edited Research Questions

Four research questions were presented in the proposed study. However, the data required to answer these questions were not available in the extant data set. As a result, the research questions were integrated and edited to reflect the available data. The proposed research questions, as well as, the edited questions are listed below.

The first two research questions were integrated and analyzed together. In the proposed study, the second question intended to use an analysis of variance (i.e. ANOVA) to examine differences between schools on the RtI measure. However, the sample size (i.e. number of schools) was too small, and there was not enough variation in the ratings for the analysis to be meaningful (Rutherford, 2000). Instead, the current study qualitatively described the differences between the schools' levels of RtI implementation on the five subscales and the total RtI scale.

The third proposed research question intended to examine if RtI implementation predicted student outcomes, using multiple regression, and the fourth proposed question intended to examine if the predictors differed by grade level. To run a multiple regression analysis, the data for the predictor variables (i.e. RtI scale & subscales) and the dependent variables (i.e. student outcomes) must be at the same level. However, the extant data set provided varied levels of data, including student-level outcomes and aggregated school-level RtI ratings. Further, the researcher could not aggregate student outcomes (i.e. use the mean) to run the analysis because the sample size (i.e. eight schools) was too small. The rule of thumb for running a multiple regression is to have a minimum of 30 data points (Harlow, 2014), and the current study would only have eight data points. Additionally, all grades levels were

not available at each school, which required the data to be analyzed by grade level. As a result, the third and fourth research questions were integrated, and modified to examine differences in student outcomes by school, and by grade level.

The proposed research questions:

1. To what extent are staff in elementary schools implementing the essential components of RtI with integrity (using an integrity rubric identified by the American Institutes for Research)?
2. To what extent are there differences between the integrity of implementation of the essential components of RtI, and overall RtI, as measured by the RtI Essential Components Worksheet between schools; and if differences exist, how can those differences be characterized?
3. To what extent does the integrity of overall RtI implementation, and the implementation of the five essential components, in elementary schools, predict school-level student outcomes, as measured by AIMSWEB screening data?
4. To what extent does the relationship between the integrity of RtI implementation, as measured by the RtI Essential Components Worksheet, and student outcomes, as measured by AIMSWEB screening data, vary by grade level?

The edited research questions were examined in the current study:

1. To what extent are staff in elementary schools implementing the essential components of RtI with integrity, as measured by the RtI Essential Components Worksheet between schools; and if differences exist between schools, how can those differences be characterized?
2. To what extent does the integrity of overall RtI implementation, and the implementation of the five essential components, in elementary schools, relate to differences student-level student outcomes, at each grade across schools, as measured by AIMSWEB screening data?

Appendix B: RtI Essential Components Worksheet

<div style="text-align: center;"><p>Center on RESPONSE to INTERVENTION <small>an American Institutes for Research</small></p><p>RTI Essential Components Worksheet</p></div> <div>School: _____ District: _____ Date: _____</div> <div>Person(s) Interviewed: _____</div> <div>Interviewer(s): _____</div> <div><p>Purpose:</p><p>The purpose of this worksheet is to provide a tool for collecting relevant information and for recording a school's rating on various items related to response to intervention (RTI) implementation. Descriptions of ratings for each item are provided on the RTI Essential Components Integrity Rubric (the Rubric).</p><p>Information about school-level implementation (Grades K-8) may be collected through interviews with school personnel and through observations and document review. After all of the information has been collected, use your notes and the Rubric to rate the school on each item. The Rubric provides a five-point rating scale and descriptions of practices that would score a 1, 3, or 5. Data collectors may assign the school a rating of 2 or 4 if the information collected suggests the school falls between the rubric descriptions. For example, if the reviewer judges a school to be performing at a level higher than the Rubric describes for a 3 rating but not quite at the level described for a 5, then the reviewer should rate the school as performing at a 4.</p></div>	<p>Center on Response to Intervention Copyright © 2014 American Institutes for Research. All rights reserved.</p> <p>RTI Essential Components Integrity Worksheet — I 2004_05/14</p>
---	--

Assessments: Screening, progress monitoring, and other supporting assessments are used to inform data-based decision making.			
Item	Sample Interview Questions	Comments/Remarks	Rating
Screening—The RTI system accurately identifies students at risk of poor learning outcomes or challenging behaviors.			
1. Screening Tools	What tools do you use for universal screening (probe across content areas)?	Notes Evidence Sources Recommendations	Rating: ① ② ③ ④ ⑤ Justification for rating
	How much attention was given to the vendor's evidence regarding the validity, reliability, and accuracy of the tools when selected?		
	Does your school have documentation from the vendor that these tools have been shown to be valid, reliable, and accurate with subgroups in your school?		
	Does staff understand how the tool is intended to be used? Can you or other staff provide evidence of the technical adequacy (i.e., reliability, validity, classification accuracy) of the tool?		

<p>2. Universal Screening</p>	<p>Describe the process for conducting screening in your school. To what extent is this process consistently followed?</p> <p>Are all students screened?</p> <p>How many times during the school year are students screened?</p> <p>Do you use a well-defined cut score or decision point to identify students at risk?</p> <p>How do you ensure that administration of screening assessments follows the developer's guidelines?</p>	<p>Notes</p> <p>Evidence Sources</p> <p>Recommendations</p>	<p>Rating: ① ② ③ ④ ⑤</p> <p>Justification of Rating</p>
<p>3. Data Points to Verify Risk</p>	<p>Do you review other information to help verify that the results of the initial screening are accurate before placing a student in secondary-level or intensive intervention? If so, what other types of assessment data do you use?</p>	<p>Notes</p> <p>Evidence Sources</p> <p>Recommendations</p>	<p>Rating: ① ② ③ ④ ⑤</p> <p>Justification of Rating</p>

<i>Progress Monitoring—Ongoing and frequent monitoring of progress quantifies rates of improvement and informs instructional practice and the development of individualized programs. Measures are appropriate for the student's grade and/or skill level.</i>			
Item	Sample Interview Questions	Comments/Remarks	Rating
Progress Monitoring Tools	What tools does your school use for progress monitoring (probe across content areas)?	Notes	Rating: ① ② ③ ④ ⑤ Justification of Rating
	Did a school or district staff consider the evidence from the vendor regarding the validity, reliability, and accuracy of the progress monitoring tool(s) when selecting it/them?	Evidence Sources	
	Does your school have documentation from the vendor that the tool(s) have been shown to be valid, reliable, and accurate with subgroups in your school? Can staff articulate the evidence supporting the rigor of the tool(s)?	Recommendations	
Progress Monitoring Process	Describe the process used for monitoring progress.	Notes	Rating: ① ② ③ ④ ⑤ Justification of Rating
	How often is the progress of students in secondary level interventions monitored?	Evidence Sources	
	How often is the progress of students in intensive intervention monitored? Does monitoring occur with sufficient frequency to show a trend in academic progress over time? How closely does administration of the progress monitoring tool(s) follow the developer's guidelines? To what extent is this process	Recommendations	

Center on Response to Intervention

RTI Essential Components Integrity Worksheet—4

	consistently followed?		
Data-Based Decision-Making—Data-based decision-making processes are used to inform instruction, movement within the multilevel system, and disability identification (in accordance with state law).			
Item	Sample Interview Questions	Comments/Remarks	Rating
Decision-Making Process	Describe how decisions are made to move students between tiers.	Notes	Rating: ① ② ③ ④ Justification of Rating
	Who is involved in decision making?		
	What data are used to inform those decisions, and how are they used?	Evidence Sources	
Data System	What criteria and guidelines are used for making decisions?		Rating: ① ② ③ ④ Justification of Rating
	To what extent are the screening, progress monitoring, and other assessment data used to inform instruction at all tiers, including the core instruction?	Recommendations	
	Is there a system for collecting and organizing student academic data, screening data, progress monitoring data, and other forms of data? If so, please describe.	Notes	
Data System	Is the system used consistently across school staff?	Evidence Sources	Rating: ① ② ③ ④ Justification of Rating
	Are instructional decisions made about students tracked in the data system or through another method (including movement between tiers)?	Recommendations	

Responsiveness to Secondary and Intensive Levels of Intervention	Describe how decisions about responsiveness to secondary-level interventions or intensive intervention are made <ul style="list-style-type: none"> • Are progress monitoring data used? • How is baseline performance established? • What goal setting method is used? (e.g., end-of-year benchmarks, rate of improvement, intra-individual framework? Are rates or norms provided by the vendor/developer? • What decision rules are used? Are the criteria implemented accurately and consistently?	Notes	Rating: ① ② ③ ④ ⑤ Justification of Rating
		<p>Evidence Sources</p> <p>Recommendations</p>	

Multilevel Instruction—The RTI framework includes a school-wide, multilevel system of instruction and interventions for preventing school failure. Commonly represented by the three-tiered triangle, multilevel instruction also is known as the multi-tiered system of support (MTSS).			
Item	Sample Interview Questions	Comments/Remarks	Rating
Primary-Level Instruction/Core Curriculum (Tier I)			
Research-Based Curriculum Materials	Describe primary-level instruction (core curriculum) materials What is the research base? When your school selected its core instructional materials, how much attention was paid to the research base?	Notes Evidence Sources Recommendations	Rating: ① ② ③ ④ Justification of Rating
Articulation of Teaching and Learning (in and across grade levels)	Describe the process that supports the articulation of teaching and learning from one grade to another. Describe the process that supports the articulation of teaching and learning among teachers in the same grade. How consistent is the learning experience among students in the same grade and subject with different teachers?	Notes Evidence Sources Recommendations	Rating: ① ② ③ ④ Justification of Rating

Differentiated Instruction	<p>To what extent do teachers in this school use student assessment data and knowledge of student readiness, language, and culture to offer students in the same class different teaching and learning strategies to address student needs?</p> <p>How consistent is this effort among the teaching staff?</p>	<p>Notes</p> <p>Evidence Sources</p> <p>Recommendations</p>	<p>Rating: ① ② ③ ④ ⑤</p> <p>Justification of Rating</p>
Standards-Based	<p>To what extent is the core curriculum in reading and mathematics aligned to state standards?</p> <p>Are the instructional materials aligned to the standards? Are model or sample lessons and activities that demonstrate effective teaching of the standards available to teachers?</p> <p>Have teachers been trained in the content of the standards and in how to use that content within their lessons?</p> <p>Are teachers utilizing their training and aligning their instruction to these standards?</p>	<p>Notes</p> <p>Evidence Sources</p> <p>Recommendations</p>	<p>Rating: ① ② ③ ④ ⑤</p> <p>Justification of Rating</p>

Exceeding Benchmark	<p>Are programs and activities provided to enrich or augment the curriculum for students exceeding benchmark? If so, please describe.</p> <p>Are any of these programs and activities available above and beyond the core instruction?</p>	<p>Notes</p> <p>Evidence Sources</p> <p>Recommendations</p>	<p>Rating: ① ② ③ ④ ⑤</p> <p>Justification of Rating</p>
Secondary-Level Intervention (Tier II)			
Evidence-Based Intervention	<p>What program(s) does your school use for secondary-level intervention?</p> <p>How were these programs selected?</p> <p>Have these programs demonstrated efficacy with the target populations (e.g., has research shown that the interventions positively impact student achievement)?</p>	<p>Notes</p> <p>Evidence Sources</p> <p>Recommendations</p>	<p>Rating: ① ② ③ ④ ⑤</p> <p>Justification of Rating</p>

Complements Core Instruction	How do instructors of secondary-level interventions ensure that the content they address is well aligned and complements the core instruction for each student?	Notes Evidence Sources Recommendations	Rating: ① ② ③ ④ ⑤ Justification of Rating
Instructional Characteristics	Are the secondary level interventions always led by staff adequately trained to implement the interventions with fidelity? If not, who provides the secondary level intervention and what is their background? Are the secondary interventions always conducted with small groups of students? What is the maximum group size?	Notes Evidence Sources Recommendations	Rating: ① ② ③ ④ ⑤ Justification of Rating
Addition to Primary	Are secondary-level interventions (i.e., Tier II) always implemented as supplements to the core curriculum? If not, please explain.	Notes Evidence Sources Recommendations	Rating: ① ② ③ ④ ⑤ Justification of Rating

<i>Intensive Intervention—Individualized with a focus on the academic needs of students with disabilities and those significantly below grade level</i> (Tier III)				
Data-Based Interventions Adapted Based on Student Need	How are evidence-based interventions intensified or individualized at the intensive level?	Notes Evidence Sources Recommendations	Rating: ① ② ③ ④ ⑤ Justification of Rating	
	How are the interventions used at this level developed?			
Instructional Characteristics	Who provides intensive intervention? Can you describe their background and level of training in providing data-based individualized instruction?	Notes Evidence Sources Recommendations	Rating: ① ② ③ ④ ⑤ Justification of Rating	
	Does the group size allow for the interventionist to adjust and individualize instruction to address the needs of each student?			
	Does an example of a student experiencing intensive intervention			

Relationship to Primary	Are intensive interventions always implemented as supplements to the core curriculum? If not, please explain. How do you decide if a student receiving intensive intervention should remain in primary-level instruction? How do you ensure meaningful connections between intensive intervention and the general education curriculum (e.g., the Common Core)?	Notes	Rating: ① ② ③ ④ ⑤ Justification of Rating
		Evidence Sources	
		Recommendations	

<i>Infrastructure and Support Mechanisms—Knowledge, resources, and organizational structures necessary to operationalize all components of RTI in a unified system to meet the established goals.</i>			
Item	Sample Interview Questions	Comments/Remarks	Rating: ① ② ③ ④ ⑤ Justification of Rating
Prevention Focus	To what extent do you believe the teaching staff views the purpose of RTI as primarily to prevent students from having academic and/or behavioral problems? What portion of the teaching staff view RTI as primarily a means for special education identification?	Notes Evidence Sources Recommendations	

Leadership Personnel	<p>To what extent are the school and district administrators aware of the RTI framework at your school?</p> <p>To what extent do the actions taken and decisions made by district administrators improve the effectiveness of the RTI framework at your school?</p> <p>To what extent do the actions taken and decisions made by school administrators improve the effectiveness of the RTI framework at your school?</p> <p>Does your school have a designated person who oversees and manages RTI implementation? If yes, what percentage of that person's time is devoted to overseeing and managing RTI?</p>	<p>Notes</p> <p>Evidence Sources</p> <p>Recommendations</p>	<p>Rating: ① ② ③ ④ ⑤</p> <p>Justification of Rating</p>
School-Based Professional Development	<p>Has the staff been trained on the RTI framework and essential components? How often is refresher or new training provided?</p> <p>Is RTI training provided to new teachers?</p> <p>What ongoing professional development is made available for those who provide secondary-level and intensive intervention?</p>	<p>Notes</p> <p>Evidence Sources</p> <p>Recommendations</p>	<p>Rating: ① ② ③ ④ ⑤</p> <p>Justification of Rating</p>

Schedules	<p>Does the schedule reflect additional time beyond the core for secondary-level and intensive intervention?</p> <p>Is there time scheduled for teacher collaboration on instruction and interventions?</p> <p>Are all the pertinent teachers and interventions available for these collaborative meetings?</p>	<p>Notes</p> <p>Evidence Sources</p> <p>Recommendations</p>	<p>Rating: ① ② ③ ④ ⑤</p> <p>Justification of Rating</p>
Resources	<p>Are there adequate materials, programs, and resources allocated to support interventions, assessments, professional development, staffing?</p> <p>Do the programs and materials match the needs of the students at each tier?</p> <p>Is there a process for monitoring the use of resources?</p>	<p>Notes</p> <p>Evidence Sources</p> <p>Recommendations</p>	<p>Rating: ① ② ③ ④ ⑤</p> <p>Justification of Rating</p>

Cultural and Linguistic Responsiveness	<p>What efforts have been made to ensure that core instruction, secondary-level and intensive intervention, and assessment is take into account cultural and linguistic factors?</p> <p>How are the demographic and academic data of subgroups represented in your school used to inform the RTI framework?</p>	<p>Notes</p> <p>Evidence Sources</p> <p>Recommendations</p>	<p>Rating: ① ② ③ ④ ⑤</p> <p>Justification of Rating</p>
Communications With and Involvement of Parents	<p>Are parents knowledgeable about the RTI framework in your school?</p> <p>Describe how you communicate with parents about RTI and student performance.</p> <p>How are parents involved in decision making regarding the participation of their child in secondary-level or intensive intervention?</p> <p>How are parents of students at the secondary or intensive level informed of the progress of their children?</p>	<p>Notes</p> <p>Evidence Sources</p> <p>Recommendations</p>	<p>Rating: ① ② ③ ④ ⑤</p> <p>Justification of Rating</p>

Communication With and Involvement of All Staff	<p>Are teachers in your school knowledgeable about the RTI framework?</p> <p>Describe how you communicate with teachers about the school's RTI plan.</p> <p>How are teachers of students at the secondary or intensive level informed of their progress in the intervention?</p> <p>What process does your school use to ensure teacher collaboration in implementing RTI?</p>	<p>Notes</p> <p>Evidence Sources</p> <p>Recommendations</p>	<p>Rating: ① ② ③ ④ ⑤</p> <p>Justification of Rating</p>
RTI Teams	<p>Does your school have an RTI team? If so:</p> <ul style="list-style-type: none"> • Who composes that team? • How often does the team meet? • Are there established processes and protocols that help the team work effectively? What are they? <p>How does the team communicate and collaborate with other staff?</p>	<p>Notes</p> <p>Evidence Sources</p> <p>Recommendations</p>	<p>Rating: ① ② ③ ④ ⑤</p> <p>Justification of Rating</p>

Fidelity and Evaluation—System for collecting and analyzing data to measure fidelity and effectiveness of the RTI model.			
Item	Sample Interview Questions	Comments/Remarks	Rating: ① ② ③ ④ Justification of Rating
Fidelity	<p>Are procedures in place to monitor the fidelity of implementation of the core curriculum? Of secondary-level and intensive intervention? Of screening, progress monitoring, and the decision-making process? If so please describe.</p> <p>Who is involved in monitoring the fidelity of implementation?</p> <p>Does the evidence indicate that instruction, interventions, assessments, and decisions are implemented with fidelity?</p>	<p>Notes</p> <p>Evidence Sources</p> <p>Recommendations</p>	<p>Rating: ① ② ③ ④ Justification of Rating</p>
Evaluation	<p>How is RTI evaluated at your school?</p> <ul style="list-style-type: none"> Is a plan in place for evaluation? Is a process in place for reviewing student-level data for all students and for subgroups of students? Is a process in place to evaluate implementation fidelity? <p>How are evaluation data used?</p> <ul style="list-style-type: none"> Are teachers and interventionists involved in giving and receiving feedback on the effectiveness of the programs and materials? <p>Who is involved in evaluating RTI implementation?</p>	<p>Notes</p> <p>Evidence Sources</p> <p>Recommendations</p>	<p>Rating: ① ② ③ ④ Justification of Rating</p>

RTI Fidelity of Implementation Rubric

The Response to Intervention (RTI) Fidelity Rubric is for use by individuals who are responsible for monitoring school-level fidelity of RTI implementation. The rubric is aligned with the essential components of RTI and the infrastructure that is necessary for successful implementation. It is accompanied by a worksheet with guiding questions and score points for use in an interview with a school's RTI leadership team.

Assessment: Screening, progress monitoring, and other supporting assessments are used to inform data-based decision making.				
Measures		1	3	5
Screening—The RTI framework accurately identifies students at risk of poor learning outcomes or challenging behaviors.				
Screening Tools	In sufficient evidence that the screening tools are reliable, correlations between the instruments and valued outcomes are strong, and predictions of risk status are accurate.	Evidence indicates that the screening tools are reliable, correlations between the instruments and valued outcomes are strong, and predictions of risk status are accurate, but staff is unable to articulate the supporting evidence.	Evidence indicates that the screening tools are reliable, correlations between the instruments and valued outcomes are strong, and predictions of risk status are accurate, and staff is able to articulate the supporting evidence.	
Universal Screening	One or none of the following conditions is met: (1) screening is conducted for all students (i.e., is universal); (2) procedures are in place to ensure implementation accuracy (i.e., all students are tested, scores are accurate, cut points/decisions are accurate); and (3) a process to screen all students occurs more than once per year (e.g., fall, winter, spring).	Two of the following conditions are met: (1) screening is conducted for all students (i.e., is universal); (2) procedures are in place to ensure implementation accuracy (i.e., all students are tested, scores are accurate, cut points/decisions are accurate); and (3) a process to screen all students occurs more than once per year (e.g., fall, winter, spring).	All of the following conditions are met: (1) screening is conducted for all students (i.e., is universal); (2) procedures are in place to ensure implementation accuracy (i.e., all students are tested, scores are accurate, cut points/decisions are accurate); and (3) a process to screen all students occurs more than once per year (e.g., fall, winter, spring).	
Data Points to Verify Risk	Screening data are not used or are used alone to verify decisions about whether a student is or is not at risk.	Screening data are used in concert with at least one other data source (e.g., classroom performance, curriculum-based assessment, performance on state-based assessment).	Screening data are used in concert with at least two other data sources (e.g., classroom performance, performance on state assessments, diagnostic assessment).	

		assessments, diagnostic assessment data, short-term progress monitoring) to verify decisions about whether a student is or is not at risk.	data, short-term progress monitoring) to verify decisions about whether a student is or is not at risk.
Progress Monitoring—Choosing and frequent monitoring of progress quantifies rates of improvement and informs instructional practice and the development of individualized programs. Measures are appropriate for the student's grade and/or skill level.			
Progress-Monitoring Tools	Selected progress-monitoring tools meet no more than one of the following criteria: (1) have sufficient number of alternate forms of equal and controlled difficulty to allow for progress monitoring at recommended intervals based on intervention level; (2) specify minimum acceptable growth; (3) provide benchmarks for minimum acceptable end-of-year performance; and (4) reliability and validity information for the performance-level score is available.	Selected progress-monitoring tools meet two or three of the following criteria: (1) have sufficient number of alternate forms of equal and controlled difficulty to allow for progress monitoring at recommended intervals based on intervention level; (2) specify minimum acceptable growth; (3) provide benchmarks for minimum acceptable end-of-year performance; and (4) reliability and validity information for the performance-level score is available.	Selected progress-monitoring tools meet all of the following criteria: (1) have sufficient number of alternate forms of equal and controlled difficulty to allow for progress monitoring at recommended intervals based on intervention level; (2) specify minimum acceptable growth; (3) provide benchmarks for minimum acceptable end-of-year performance; and (4) reliability and validity information for the performance-level score is available and staff is able to articulate the supporting evidence.
Progress-Monitoring Process	Neither of the following conditions is met: (1) progress monitoring occurs at least monthly for students receiving secondary-level intervention and at least weekly for students receiving intensive intervention; and (2) procedures are in place to ensure implementation accuracy (i.e., appropriate students are tested, scores are accurate, decision-making rules are applied consistently).	Only one of the following conditions is met: (1) progress monitoring occurs at least monthly for students receiving secondary-level intervention and at least weekly for students receiving intensive intervention; and (2) procedures are in place to ensure implementation accuracy (i.e., appropriate students are tested, scores are accurate, decision-making rules are applied consistently).	Both of the following conditions are met: (1) progress monitoring occurs at least monthly for students receiving secondary-level intervention and at least weekly for students receiving intensive intervention; and (2) procedures are in place to ensure implementation accuracy (i.e., appropriate students are tested, scores are accurate, decision-making rules are applied consistently).

Center on Response to Intervention

RTI Fidelity of Implementation Rubric—2

Data-Based Decision Making—Data-based decision-making processes are used to inform instruction, movement within the multi-level system, and disability identification (in accordance with state law).				
Measures		1	3	5
Decision-Making Process		The mechanism for making decisions about the participation of students in the instruction/intervention levels meets more than one of the following criteria: The process (1) is data-driven and based on validated methods; (2) involves a broad base of stakeholders; and (3) is operationalized with clear, established decision rules (e.g., movement between levels or tiers, determination of appropriate instruction or interventions).	The mechanism for making decisions about the participation of students in the instruction/intervention levels meets two of the following criteria: The process (1) is data-driven and based on validated methods; (2) involves a broad base of stakeholders; and (3) is operationalized with clear, established decision rules (e.g., movement between levels or tiers, determination of appropriate instruction or interventions).	The mechanism for making decisions about the participation of students in the instruction/intervention levels meets all of the following criteria: The process (1) is data-driven and based on validated methods; (2) involves a broad base of stakeholders; and (3) is operationalized with clear, established decision rules (e.g., movement between levels or tiers, determination of appropriate instruction or interventions).
Data System		A data system is in place that meets two or fewer of the following conditions: (1) the system allows users to document and access individual student-level data (including screening and progress-monitoring data) and instructional decisions; (2) data are entered in a timely manner; (3) data can be represented graphically; and (4) there is a process for setting/evaluating goals.	A data system is in place that meets three of the following four conditions: (1) the system allows users to document and access individual student-level data (including screening and progress-monitoring data) and instructional decisions; (2) data are entered in a timely manner; (3) data can be represented graphically; and (4) there is a process for setting/evaluating goals.	A data system is in place that meets all of the following conditions: (1) the system allows users to document and access individual student-level data (including screening and progress-monitoring data) and instructional decisions; (2) data are entered in a timely manner; (3) data can be represented graphically; and (4) there is a process for setting/evaluating goals.
Responsiveness to Secondary and Intensive Levels of Intervention		Neither of the following conditions is met: (1) decisions about responsiveness to intervention are based on reliable and valid progress-monitoring data that reflect slope of improvement or progress toward the attainment of a goal at the end of the intervention; and (2) these decision-making criteria are implemented accurately.	Only one of the following conditions is met: (1) decisions about responsiveness to intervention are based on reliable and valid progress-monitoring data that reflect slope of improvement or progress toward the attainment of a goal at the end of the intervention; and (2) these decision-making criteria are implemented accurately.	Both of the following conditions are met: (1) decisions about responsiveness to intervention are based on reliable and valid progress-monitoring data that reflect slope of improvement or progress toward the attainment of a goal at the end of the intervention; and (2) these decision-making criteria are implemented accurately.

Center on Response to Intervention

RTI Fidelity of Implementation Rubric—3

Multilevel Instruction—The RTI framework includes a school-wide, multi-level system of instruction and interventions for preventing school failure. Commonly represented by the three-tiered triangle, multilevel instruction also is known as the multi-tiered system of support (MTSS).				
Measures	1	2	3	5
Primary-Level Instruction Core Curriculum (Tier I)				
Research-Based Curriculum Materials	Few core curriculum materials are research based for the target population of learners (including subgroups).	Some core curriculum materials are research based for the target population of learners (including subgroups).	All core curriculum materials are research based for the target population of learners (including subgroups).	
Articulation of Teaching and Learning (in and across grade levels)	Neither of the following conditions is met: (1) teaching and learning objectives are well articulated from one grade to another, and (2) teaching and learning is well articulated within grade levels so that students have highly similar experiences, regardless of their assigned teacher.	Only one of the following conditions is met: (1) teaching and learning objectives are well articulated from one grade to another, and (2) teaching and learning is well articulated within grade levels so that students have highly similar experiences, regardless of their assigned teacher.	Both of the following conditions are met: (1) teaching and learning objectives are well articulated from one grade to another, and (2) teaching and learning is well articulated within grade levels so that students have highly similar experiences, regardless of their assigned teacher.	
Differentiated Instruction	Neither of the following conditions is met: (1) interviewed staff can describe how most teachers in the school differentiate instruction for students on, below, or above grade level, and (2) interviewed staff can explain how most teachers in the school use student data to identify and address the needs of students.	Only one of the following conditions is met: (1) interviewed staff can describe how most teachers in the school differentiate instruction for students on, below, or above grade level, and (2) interviewed staff can explain how most teachers in the school use student data to identify and address the needs of students.	Both of the following conditions are met: (1) interviewed staff can describe how most teachers in the school differentiate instruction for students on, below, or above grade level, and (2) interviewed staff can explain how most teachers in the school use data to identify and address the needs of students.	
Standards-Based	The core curriculum (reading and mathematics) is not aligned with the Common Core or other state standards.	The core curriculum (reading and mathematics) is partially aligned with the Common Core or other state standards.	The core curriculum (reading and mathematics) is aligned with the Common Core or other state standards.	

Center on Response to Intervention

RTI Fidelity of Implementation Rubric—4

Exceeding Benchmark	Neither of the following conditions is met: (1) the school provides enrichment opportunities for students exceeding benchmarks, and (2) teachers implement those opportunities consistently at all grade levels.	One of the following conditions is met: (1) the school provides enrichment opportunities for students exceeding benchmarks, and (2) teachers implement those opportunities consistently at all grade levels.	Both of the following conditions are met: (1) the school provides enrichment opportunities for students exceeding benchmarks, and (2) teachers implement those opportunities consistently at all grade levels.
Secondary-Level Intervention (Tier II)			
Evidence-Based Intervention	Secondary-level interventions are not evidence based in content areas and grade levels where they are available.	Some secondary-level interventions are evidence based in content areas and grade levels where they are available.	All secondary-level interventions are evidence based in content areas and grade levels where they are available.
Complements Core Instruction	Secondary-level intervention is poorly aligned with core instruction and incorporates different topics, even though those topics are not foundational skills that support core program learning objectives.	Secondary-level intervention incorporates foundational skills, but these only occasionally align with the learning objectives of core instruction.	Secondary-level intervention is well aligned with core instruction and incorporates foundational skills that support the learning objectives of core instruction.
Instructional Characteristics	One or none of the following conditions is met: (1) interventions are standardized, (2) secondary-level interventions are led by staff trained in the intervention according to developer requirements, and (3) group size and dosage are optimal (according to research) for the age and needs of students.	Two of the following conditions are met: (1) interventions are standardized, (2) secondary-level interventions are led by staff trained in the intervention according to developer requirements, and (3) group size and dosage are optimal (according to research) for the age and needs of students.	All three of the following conditions are met: (1) interventions are standardized, (2) secondary-level interventions are led by staff trained in the intervention according to developer requirements, and (3) group size and dosage are optimal (according to research) for the age and needs of students.
Addition to Primary	Secondary-level interventions replace core instruction.	Secondary-level interventions sometimes supplement core instruction and sometimes replace core instruction.	Secondary-level interventions supplement core instruction.

<i>Intensive Intervention—Individualized with a focus on the academic needs of students with disabilities and those significantly below grade level (Tier III)</i>				
Data-Based Interventions Adapted Based on Student Need	Intensive interventions are not more intensive (e.g., no increase in duration or frequency, change in interventionist, change in group size, or change in intervention) than secondary interventions.	Intensive interventions are more intensive than secondary interventions based only on preset methods to increase intensity (e.g., sole reliance on increased duration or frequency, change in interventionist, decreased group size, or change in intervention program).	Intensive interventions are more intensive than secondary interventions and are adapted to address individual student needs in a number of ways (e.g., increased duration or frequency, change in interventionist, decreased group size, change in instructional delivery, and change in type of intervention) through an iterative manner based on student data.	
Instructional Characteristics	None of the following conditions is met: (1) the intervention is individualized; (2) intensive interventions are led by well-trained staff experienced in individualizing instruction based on student data; and (3) the group size is optimal (according to research) for the age and needs of students.	Only one of the following conditions is met: (1) the intervention is individualized; (2) intensive interventions are led by well-trained staff experienced in individualizing instruction based on student data; and (3) the group size is optimal (according to research) for the age and needs of students.	All of the following conditions are met: (1) the intervention is individualized; (2) intensive interventions are led by well-trained staff experienced in individualizing instruction based on student data; and (3) the group size is optimal (according to research) for the age and needs of students.	
Relationship to Primary	Neither of the following conditions is met: (1) decisions regarding student participation in both core instruction and intensive intervention are made on a case-by-case basis, according to student need; and (2) intensive interventions are aligned to the specific skill needs of students to help them make progress toward core curriculum standards.	Only one of the following conditions is met: (1) decisions regarding student participation in both core instruction and intensive intervention are made on a case-by-case basis, according to student need; and (2) intensive interventions address the general education curriculum in an appropriate manner for students.	Both of the following conditions are met: (1) decisions regarding student participation in both core instruction and intensive intervention are made on a case-by-case basis, according to student need; and (2) intensive interventions address the general education curriculum in an appropriate manner for students.	

Infrastructure and Support Mechanisms—Knowledge, resources, and organizational structures necessary to operationalize all components of RTI in a unified system to meet the established goals.				
Measures	1	3	5	
Prevention Focus	Staff generally perceives RTI as a program that solely supports the preferential process for special education.	Some staff understand that RTI is a framework to prevent all students, including students with disabilities, from having academic problems.	All staff understand that RTI is a framework to prevent all students, including students with disabilities, from having academic problems.	
Leadership Personnel	Decisions and actions by school and district leaders undermine the effectiveness of the essential components of the RTI framework at the school.	Decisions and actions by school and district leaders are inconsistent and only somewhat supportive of the essential components of the RTI framework at the school; support for RTI implementation is not very evident.	Decisions and actions by school and district leaders proactively support the essential components of the RTI framework at the school, and help make the RTI framework more effective; support for RTI implementation is a high priority.	
School-Based Professional Development	The school has no well-defined, school-based professional development mechanisms to support continuous improvement of instructional practice, data-based decision making, and delivery of interventions.	Some forms of school-based professional development are available, but most are not consistent or job embedded to ensure continuous improvement in instructional practice, data-based decision making, and delivery of interventions.	School-based professional development is institutionalized and structured so that all teachers continuously examine, reflect upon, and improve instructional practice, data-based decision making, and delivery of interventions.	
Schedules	School wide schedules are not aligned to support multiple levels of intervention based on student need; inadequate time is available for interventions.	School wide schedules are partially aligned to support multiple levels of intervention based on student need; some additional time is built in for interventions.	School wide schedules are aligned to support multiple levels of intervention based on student need; adequate additional time is built in for interventions.	
Resources	Resources (e.g., funds, programs) are not allocated to support RTI implementation.	Resources (e.g., funds, programs) are partially allocated to support RTI implementation.	Resources (e.g., funds, programs) are adequately allocated to support RTI implementation.	
Cultural and Linguistic Responsiveness	One or none of the following conditions is met: Staff can articulate information and factors that they consider when adopting	Two of the following conditions are met: Staff can articulate information and factors that they consider when adopting	All three of the following conditions are met: Staff can articulate information and factors that they consider when adopting	

Center on Response to Intervention

RTI Fidelity of Implementation Rubric—7

	culturally and linguistically relevant (1) instructional practices, (2) assessments, and (3) intervention programs.	culturally and linguistically relevant (1) instructional practices, (2) assessments, and (3) intervention programs.	culturally and linguistically relevant (1) instructional practices, (2) assessments, and (3) intervention programs.
Communications With and Involvement of Parents	One or none of the following conditions is met: (1) a description of the school's essential components of RTI is shared with parents; (2) a coherent mechanism is implemented for updating parents on the progress of their child who is receiving secondary or intensive interventions; and (3) parents are involved during decision making regarding the progress of students receiving intensive intervention.	Two of the following conditions are met: (1) a description of the school's essential components of RTI is shared with parents; (2) a coherent mechanism is implemented for updating parents on the progress of their child who is receiving secondary or intensive interventions; and (3) parents are involved during decision making regarding the progress of students receiving intensive intervention.	All of the following conditions are met: (1) a description of the school's essential components of RTI is shared with parents; (2) a coherent mechanism is implemented for updating parents on the progress of their child who is receiving secondary or intensive interventions; and (3) parents are informed about decision making regarding the progress of students receiving intensive intervention.
Communication With and Involvement of All Staff	One or none of the following conditions is met: (1) a description of the school's essential components of RTI and data-based decision-making process is shared with staff; (2) a system is in place to keep staff informed; and (3) teacher teams collaborate frequently.	At least two of the following conditions are met: (1) a description of the school's essential components of RTI and data-based decision-making process is shared with staff; (2) a system is in place to keep staff informed; and (3) teacher teams collaborate frequently.	All of the following conditions are met: (1) a description of the school's essential components of RTI and data-based decision-making process is shared with staff; (2) a system is in place to keep staff informed; and (3) teacher teams collaborate frequently.
RTI Teams	Only one of the following conditions is met: (1) the RTI team is representative of all key stakeholders; (2) structures and clear processes are in place to guide decision making; and (3) time is set aside for the team to meet regularly.	At least two of the following conditions are met: (1) the RTI team is representative of all key stakeholders; (2) structures and clear processes are in place to guide decision making; and (3) time is set aside for the team to meet regularly.	All of the following conditions are met: (1) the RTI team is representative of all key stakeholders; (2) structures and clear processes are in place to guide decision making; and (3) time is set aside for the team to meet regularly.

Fidelity and Evaluation—System for collecting and analyzing data to measure fidelity and effectiveness of the RTI model.				
Measures	1	3	5	
Fidelity	Neither of the following conditions is met: (1) procedures are in place to monitor the fidelity of implementation of the core curriculum and secondary and intensive interventions, and (2) procedures are in place to monitor the processes of administering and analyzing assessments.	One of the following conditions is met: (1) procedures are in place to monitor the fidelity of implementation of the core curriculum and secondary and intensive interventions, and (2) procedures are in place to monitor the processes of administering and analyzing assessments.	Both of the following conditions are met: (1) procedures are in place to monitor the fidelity of implementation of the core curriculum and secondary and intensive interventions, and (2) procedures are in place to monitor the processes of administering and analyzing assessments.	
Evaluation	None of the following conditions are met: (1) an evaluation plan is in place to monitor short- and long-term goals, (2) student data are reviewed for all students and subgroups of students across the essential components to evaluate effectiveness of the RTI framework (i.e., core curriculum is effective, interventions are effective, screening process is effective), and (3) implementation data (e.g., walk-throughs) are reviewed to monitor fidelity and efficiency across all components of the RTI framework.	At least one of the following conditions is met: (1) an evaluation plan is in place to monitor short- and long-term goals, (2) student data are reviewed for all students and subgroups of students across the essential components to evaluate effectiveness of the RTI framework (i.e., core curriculum is effective, interventions are effective, screening process is effective), and (3) implementation data (e.g., walk-throughs) are reviewed to monitor fidelity and efficiency across all components of the RTI framework.	All of the following conditions are met: (1) an evaluation plan is in place to monitor short- and long-term goals, (2) student data are reviewed for all students and subgroups of students across the essential components to evaluate effectiveness of the RTI framework (i.e., core curriculum is effective, interventions are effective, screening process is effective), and (3) implementation data (e.g., walk-throughs) are reviewed to monitor fidelity and efficiency across all components of the RTI framework.	

Appendix D: RtI Measure Reliability

Cronbach's alpha (i.e. r_α), a measure of internal consistency, was used to estimate the reliability of the RtI measure. The RtI fidelity of implementation worksheet and rubric were designed to measure the five essential components of RtI implementation, as well as, to provide one overall measure of RtI implementation. In other words, the measure was designed to be multi-dimensional. Accordingly, Cronbach's alpha was calculated for each of the five essential components, and for total RtI, to determine if the items were measuring the same underlying construct (Pallant, 2005). Internal consistency estimates are dependent on the sample size and the number of items contributing to a scale or subscale (Tavakol & Dennick, 2011). The recommended sample size for calculating coefficient alpha is 50 (Javali & Gudaganavar, 2011). Further, sample size may affect Cronbach's alpha as a result of limited variation in the scores. The analysis for the current study was limited because of the small sample size ($N=8$) and the small number of items on two subscales (i.e. 2 items on DBDM & fidelity and evaluation).

Limited information about scale reliability was available from prior studies that used the RtI fidelity of implementation measure. In the current study, first, the researcher examined the internal consistency of the measure using the sample of eight elementary schools. The results are presented in the table below. A negative coefficient was found for the subscale 'assessment' ($r_\alpha = -.774$), which indicates that there was not enough variation in the items to conduct the coefficient alpha. Low reliability was found for DBDM ($r_\alpha = .615$); however, the small number of items likely affected the analysis (i.e. 2 items because one item was removed due to zero variance).

The MTSS ($r_\alpha = .843$) and infrastructure and support ($r_\alpha = .811$) subscales fell in the good reliability range (Nunnally, 1967). Cronbach's alpha was not calculated for the fidelity and evaluation subscale because one of the two items was removed due to zero variance. Finally, Cronbach's alpha was calculated for total RtI, which included all 29 items, and fell in the acceptable reliability range ($r_\alpha = .776$).

Given the limitations in the first analysis, the researcher additionally examined the internal consistency of the measure using a sample of 15 schools, which included eight elementary schools, two middle schools, and six high schools, as a comparison with a larger sample and greater variation. The results are presented in the table below. When all 15 schools were examined together, and there was greater variation in the ratings, the results started to approach more acceptable levels of reliability across all subscales. Although the internal consistency analysis was limited given the small sample size and the subscales with too few items, it is encouraging that in both samples, the total RtI fell in the high reliability range. The internal consistency of the RtI measure should be examined in future research with a larger sample size.

RtI Scale Reliability				
	8 Elementary Schools		15 Schools	
	N of items	Cronbach's Alpha	N of items	Cronbach's Alpha
Assessment	3	.774	5	.623
DBDM	2*	.615	2*	.735
MTSS	12	.843	12*	.849
Infrastructure & Support	7	.811	7	.844
Fidelity & Evaluation		N/A**	2	.157
Total RtI (29 items)	25	.776	26	.925

*one item was removed due to zero variance

**only 2 items and 1 has zero variance so cannot calculate

Appendix E: Normality Information and Comparison of Parametric and Nonparametric Test Results

Part 1: Histograms and Normal Q-Q Plots

Below are visual representations (i.e. histogram & normal Q-Q plots) of the distribution, by school, for each of the five variables that did not meet the assumption of normality. It is important to note that all variables met the requirement for skewness and kurtosis. However, further investigation including the Shapiro-Wilk test and normal Q-Q plots indicated departures from normality. In normal Q-Q plots, a reasonably straight line suggests a normal distribution, and deviations from a reasonably straight line indicate departures from normality (Pallant, 2005). An analysis of the normal Q-Q plots below suggests that some variables moderately failed the normality assumption as a result of outliers. Consequently, the nonparametric test equivalent to the ANOVA, the Kruskal-Wallis test, was conducted. The nonparametric equivalent may be used because it does not require the assumption of normality (Harlow, 2014).

Grade 1 ROI

Grade 3 ROI

Grade 4 ROI

Grade 5 ROI

Part B: Justification for Selected Analysis

The table below presents a comparison of results from the parametric test (i.e. ANOVA) and the nonparametric equivalent test (i.e. Kruskal-Wallis test), for all variables that did not meet the normality assumption. Results from the Kruskal-Wallis test were consistent with the results from the ANOVA on all variables except grade 1 WRC and grade 4 ROI. Given that the skewness and kurtosis were acceptable, only some of the normal Q-Q plots moderately failed due to outliers, and the results for the two analyses were consistent, the researcher chose to report the ANOVA results for all variables except for grade 1 WRC and grade 4 ROI. In summary, grade 1 WRC and grade 4 ROI were analyzed using the Kruskal-Wallis test, and the remaining variables listed below were analyzed using a one-way ANOVA.

A Comparison of Parametric results (ANOVA) vs. Nonparametric results (Kruskal-Wallis Test) for dependent variables with non-normal distributions				
	ANOVA		Kruskal-Wallis Test	
	F	Sig.	Chi-Square	Asymp. Sig.
Grade 1 WRC	2.254	.037*	10.507	.105
Grade 1 ROI	6.673	.000**	34.395	.000**
Grade 3 ROI	5.169	.000**	28.929	.000**
Grade 4 ROI	4.18	.62	20.871	.001**
Grade 5 ROI	4.757	.003**	13.788	.003**

* Significant at the .05 level

** Significant at the .01 level

Bibliography

- American Psychological Association. (2005). *Policy statement on evidence-based practice in psychology*. Washington, DC: Author. Retrieved on September 19, 2012, from <http://www.apa.org/practice/resources/evidence/index.aspx>
- Ardoyn, 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 or oral reading fluency (CBM-R) decision rules. *Journal of School Psychology, 51*(1), 1-18.
- Bailey, T. R. (2014). An initial exploratory analysis of RTI implementation in rural schools. *The Researcher, 26*(1), 34-39.
- Berkeley, S., Bender, W. N., Peaster, L.G., & Saunders, L. (2009). Implementation of response to intervention. *Journal of Learning Disabilities, 42*(1), 85-95.
- Bollman, K.A., Silberglitt, B., & Gibbons, K. A. (2007). The St. Croix River Education district model: Incorporating systems-level organization and a multi-tiered problem-solving process for intervention delivery. In S. R. Jimerson, M. K. Burns & A. VanDerHyden (Eds.), *Handbook of Response to Intervention: The Science and Practice of Assessment and Intervention* (319-330). New York, NY: Springer.
- Brown-Chidsey, R. (2007). No more "Waiting to fail." *Educational Leadership, 65*(2), 40-46.
- Burns, M. K., Appleton, J.J., & Stehouwer, J. D. (2005). Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment, 23*, 381-394.
- Burns, M. K., and Gibbons, K. (2008). *Implementing response-to-intervention in elementary and secondary schools*. New York, NY: Taylor & Francis Group, LLC.
- Burns, M. K. & Symington, T. (2002). A meta-analysis of prereferral intervention teams: Student and systemic outcomes. *Journal of School Psychology, 40*(5), 437-447.
- Codding, R. S., Feinberg, A. B., & Dunn, E. K., & Pace, G. M. (2005). Effects of immediate performance feedback on implementation of behavior support plans. *Journal of Applied Behavior Analysis, 38*(2), 205-219.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghtin.

- Deeney, T. A., & Shim, M. K. (2016). Teachers' and students' views of reading fluency: Issues of consequential validity in adopting one minute reading fluency assessments. *Assessment for Effective Interventions*, 41(2), 109-126.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52 (3), 219-232.
- Dexter, D. D., Hughes, C. A., & Farmer, T. W. (2008). Responsiveness to intervention: A review of field studies and implications for rural special education. *Rural Special Education Quarterly*, 26(4), 3-9.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327-350.
- Fuchs, D., Fuchs, L.S., Mathes, P. G., & Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, 34(1), 174-206.
- Fuchs, L.S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review*, 22, 27-48.
- Fuchs, L. S. (2007). NCRLD update on responsiveness to intervention: Research to practice. *National Research Center on Learning Disabilities*.
- Glover, T. A., & DiPerna, J.C. (2007). Service delivery for response to intervention: Core components and directions for future research. *School Psychology Review*, 36(4), 526-540.
- Green, S.K., & Johnson, R. L (2010). *Assessment is essential*. New York, NY: McGraw Hill.
- Greenwood, C. R. (2009). Treatment integrity: Revisiting some big ideas. *School Psychology Review*, 38(4), 547.
- Greenwood, C. R., & Kim, J. M. (2012). Response to intervention (RTI) services: An ecobehavioral perspective. *Journal of Educational and Psychological Consultation*, 22, 79-105.
- Good, R. H., & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurements*. New York, NY: Guildford Press.

- Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2009). Treatment integrity assessment in the schools: An evaluation of the treatment integrity planning protocol. *School Psychology Quarterly*, 24, 24-35.
- Harlow, L. L. (2005). *The Essence of Multivariate Thinking: Basic Themes and Methods*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hill, D. R., King, A. A., Lemons, C. J., & Partanen, J. N. (2012). Fidelity if implementation and instructional alignment in response to intervention research. *Learning Disabilities Research & Practice*, 27(3), 116-124.
- Hosp, J. & Suchey, N. (2014). Reading assessment: Reading fluency, reading fluently, and comprehension. *School Psychology Review*, 43(1), 59-68.
- Hoover, J. J., Baca, L., Wexler-Love, E., & Saenz, L. (2008). National implementation of response to intervention (RTI): Research summary. Retrieved from <http://www.nasdse.org/Portals/0/NationalImplementationofRTI-ResearchSummary.pdf>
- Individuals with Disabilities Education Improvement Act (2004). P.L. No. 108-446 614 [b][6][A]; 614 [b][2&3].
- Javali, S. B., Gudaganavar, N. V., & Raj, S. M. (2011). Effect of varying sample size in estimation of coefficients of internal consistency. Retrieved from https://www.webmedcentral.com/wmcpdf/Article_WMC001649.pdf
- Jimerson, S. R., Burns, S. R., & VanDerHeyden, A. M. (Eds.). (2016). *Handbook of response to intervention: The science and practice of multi-tiered systems of support*. New York, NY: Springer US.
- Joseph, L. M. (2014). Best practices on interventions for students with reading problems. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Student-level services*, (pp. 97-114). Bethesda, MD: National Association of School Psychologists.
- Keller-Margulis, M.A. (2012). Fidelity of implementation framework: A critical need for response to intervention models. *Psychology in the Schools*, 49(4), 342-352.
- Kovaleski, J. F. (2007). Response to intervention: Considerations for research and systems change. *School Psychology Review*, 36(4), 638-646.
- Kratochwill, T. R., Volpiansky, P., Clements, M. & Ball, C. (2007). Professional development in implementing and sustaining multitier prevention models:

- Implications for response to intervention. *School Psychology Review*, 36(4), 619-631.
- Leon, Y., Wilder, D. A., Majdalany, L., Myers, K., & Saini, V. (2014). Errors of omission and commission during alternative reinforcement of compliance: The effects of varying levels of treatment integrity. *Journal of Behavioral Education*, 23, 19-33.
- Martinez, R. & Young, A. (2011). Response to intervention: How is it practiced and perceived? *International Journal of Special Education*, 26, 44-52.
- Maltzman, S. (2016). *Treatment processes and outcomes in psychology: A multidisciplinary, biopsychosocial approach*. Retrieved from https://books.google.com/books?id=csOSDAAAQBAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
- National Center for Response to Intervention. (2014). RTI framework integrity rubric. Retrieved from <http://www.rti4success.org/resource/essential-components-rti-integrity-rubric-and-worksheet>
- National Center for Response to Intervention. (2014). RTI fidelity of implementation worksheet. Retrieved from <http://www.rti4success.org/resource/essential-components-rti-integrity-rubric-and-worksheet>.
- Nellis, L. M. (2012). Maximizing the effectiveness of building teams in response to intervention implementation. *Psychology in the Schools*, 49(3), 245-256.
- Newton, J. S., Horner, R. H., Todd, A. W., Algozzine, R. F., & Algozzine, K. M. (2012). A pilot study of a problem solving model for team decision making. *Education and Treatment of Children*, 35(1), 25-29.
- No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).
- Noell, G. H., & Gansle, K. A. (2006). Assuring the form has substance. *Assessment for Effective Intervention*, 32(1), 32-39
- Noell, G. H., & Gansle, K. A. (2014). Assuring the response to intervention process has substance: Assessing and supporting intervention implementation. In S. R. Jimerson, S. R. Burns, & A. M. VanDerHeyden, (Eds.), *Handbook of response to intervention: The science and practice of multi-tiered systems of support*. (pp. 407-420). New York, NY: Springer US
- Noltemeyer, A., & Sansosti, F. J. (2012). Tiered models of integrated academic and behavioral support: Effect of implementation level on academic outcomes. *Contemporary School Psychology*, 16, 117-127.

- Noltemeyer, A. L., Boone, W. J., & Sansosti, F. J. (2016). Assessing school-level RTI implementation for reading: Development and piloting of RTIS-R. *Assessment for Effective Intervention*, 40(1), 40-52.
- Nunnally, J. C. (1967). *Psychometric Theory*. New York, NY: McGraw-Hill
- O'Connor, E. P., & Freeman, E. W. (2012). District-level considerations in supporting and sustaining RTI implementation. *Psychology in the Schools*, 49(3), 297-310
- Odom, S. L. (2009). The tie that binds: Evidence-based practice, implementation science, and outcomes for children. *Topics in Early Childhood Special Education*, 29(1), 53-61.
- Odom, S. L., Fleming, K., Diamond, K., Lieber, J., Hanson, M., Butera, G., Horn, E., Palmer, S., Marquis, J., & Children's School Success Project. (2010). Examining different forms of implementation and in early childhood curriculum research. *Early Child Research Quarterly*, 25(3), 314-328.
- Pallant, J. (2010). *SPSS Survival Manual*. Open University Press.
- 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.
- Sanetti, L. M., & Kratochwill, T. R. (Eds.). (2014). *Treatment integrity: A foundation for evidence-based practice in applied psychology*. Washington, DC: American Psychological Association.
- Schulte, A. C., Eaton, J. E., & Parker, J. (2009). Advances in treatment integrity research: Multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treatment integrity. *School Psychology Review*, 38, 460-475.
- Sharp, K., Sanders, K., Noltemeyer, A., Hoffman, J., & Boone, W. J. (2016). The relationship between RTI implementation and reading achievement: A school-level analysis. *Preventing School Failure: Alternative Education for Children and Youth*, 60(2), 152-160.
- Shapiro, E. S., & Clemens, N. H. (2009). A conceptual model for evaluating system effects of response to intervention. *Assessment for Effective Intervention*, 35(1), 3-16.
- Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, 42(8), 795-819.

- Stobier, K. C. (2014). A comprehensive framework for multitiered systems of support in school psychology. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Data-based and collaborative decision making*, (pp. 129-142). Bethesda, MD: National Association of School Psychologists.
- Swindlehurst, K., Shepard, K., Salembier, G., & Hurley, S. (2015). Implementing response to intervention: Results of a survey of school principals. *Rural Special Education Quarterly*, 34(2), 9-16.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's Alpha. *International Journal of Medical Education*, 2, 53-55.
- VanDerHyden, 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.
- Wilson VanVoorhis, C. R., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, 3(2), 43-50.