EXAMINATION OF A RATING SCALE TO ASSESS TEACHERS’ TREATMENT
ACCEPTABILITY OF READING INTERVENTIONS FOR STRUGGLING READERS
IN ELEMENTARY SCHOOLS

A dissertation submitted to the
Kent State University College
of Education, Health, and Human Services
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

By
Neslihan Unluol Unal

May 2018
A dissertation written by
Neslihan Unluol Unal
B.A., Abant Izzet Baysal University, 2008
M.A., The University of Texas at San Antonio, 2012
Ph.D., Kent State University, 2018

Approved by

________________________, Co-director, Doctoral Dissertation Committee
Andrew L. Wiley
________________________, Co-director, Doctoral Dissertation Committee
Brian R. Barber
________________________, Member, Doctoral Dissertation Committee
Aryn C. Karpinski

Accepted by

________________________, Director, School of Lifespan Development and
Mary Dellmann-Jenkins Educational Sciences
________________________, Dean, College of Education, Health, and Human
James Hannon Services

iii
Teachers’ judgements on effective reading interventions could influence the implementation of those interventions. To date, the primary focus of treatment acceptability research has been examining the treatment acceptability of behavioral interventions; however, a limited number of studies have investigated the treatment acceptability of reading interventions. Additionally, no validated measures have been developed to assess the treatment acceptability of reading interventions. Therefore, the first purpose of this study was to develop a new instrument called the Reading Intervention Rating Profile (RIRP) which was specifically designed to assess the acceptability of reading intervention procedures. Also, the second major purpose was to examine teacher factors (i.e., teaching experience, level of education, and certification type) that influence the acceptability ratings of reading interventions, as measured by the RIRP.

Participants who were 59 special education teachers and 78 general education teachers completed the 15-item RIRP along with demographic questions. Results suggested that the RIRP has sufficient internal consistency and reliability to be of use in assessments of the acceptability of reading interventions. Regarding the second objective of the study, results for both groups of teachers indicated that teachers’ acceptability
ratings of reading interventions differ based on their years of teaching experience and their certification type. However, no significant difference was found for the acceptability ratings for both groups of teachers on the RIRP based on their level of education. Finally, the limitations of the study, the implications of the research findings, and the recommendations for future research were discussed.
DEDICATION

I dedicate this dissertation to my beloved parents, Hatice and Kazim,
    My lovely husband, Mustafa, and
    My beautiful daughter, Erva Melis.
ACKNOWLEDGEMENTS

I would like to thank everyone who has provided me with support and encouragement through this entire process.

First and the foremost, I am deeply grateful to my advisor Dr. Andrew L. Wiley, who guided me patiently through at every step of this dissertation. None of this would have been possible without his vast knowledge, enthusiasm, and great encouragement. Throughout my study, Dr. Wiley provided endless support as he discussed my work with me, corrected my drafts, helped me contact with schools, gave invaluable feedback, and improved my research as a result. Dr. Wiley always guided me to plan ahead not only for my dissertation but also for my long-term academic career plans. He provided opportunities to me to attend national and international seminars and conferences and to improve my teaching ability as well as my academic skills. His supports were not only limited to guidance for my dissertation and academic development. Dr. Wiley has also opened his door to me whenever I need help and advice personally. I offer sincerest thanks for all of your support.

I would also like to gratefully thank my co-advisor Dr. Brian R. Barber, who critically read my work and provided suggestions. Dr. Barber was also generous in allowing me to improve my academic skills by providing opportunities to attend conferences, be a reviewer in academic journals, teach at university level, and collaborate with fellow doctoral students. He profoundly understands the necessities and requirements of a successful academic career and advised me wisely to work towards building a strong background to get ready for my future.
Special thank goes to my PhD committee member Dr. Aryn C. Karpinski for her invaluable assistance, expertise, and suggestions which enabled me to improve this dissertation. Without her help, feedback, and consultation, my dissertation could not advance to this degree.

My dear husband Mustafa deserves a special thank you, as he always believed me, gave me the motivation and stood beside me throughout my academic path. I would also like to thank my mother and mother-in-law who assisted me by taking care of my toddler for the countless hours while I was working on my dissertation writing. I would also like to thank my other family members, my father and younger brother. They were always supporting me and encouraging me with their best wishes. Without their help and encouragement, I would not have been here.

Finally, I would like to thank to the Republic of Turkey, Ministry of National Education for providing me the scholarship which covered my all tuitions and living expenses during my graduate education in the United States. Without this scholarship, this would not have been possible.
TABLE OF CONTENTS

DEDICATION.............................................................................................................................. vi

ACKNOWLEDGEMENTS............................................................................................................. vii

LIST OF FIGURES .................................................................................................................. xii

LIST OF TABLES ...................................................................................................................... xiii

CHAPTER

I. REVIEW OF THE LITERATURE ............................................................................................................. 1

   Introduction to the Study ........................................................................................................... 1

   Prevalence, Types, and Impact of Reading Difficulties in Elementary Grades .... 2

      Prevalence and types of reading problems ................................................................. 2

      Impact of reading on educational and developmental outcomes .................... 5

   Effective Reading Interventions in Special Education ..................................................... 7

   Research to Practice Gap .................................................................................................... 12

   Social Validity ...................................................................................................................... 16

   Treatment Acceptability ..................................................................................................... 18

      Conceptual Models of Treatment Acceptability ..................................................... 19

         Witt and Elliott's (1985) model ........................................................................ 19


         Elliott's (1988) model ......................................................................................... 22

   Synthesis of Research on Treatment Acceptability .......................................................... 25

   Factors Influencing Treatment Acceptability ............................................................... 26

      Student factors .............................................................................................................. 26

      Teacher factors ............................................................................................................ 27

      Intervention factors ..................................................................................................... 29

      Contextual factors ...................................................................................................... 32

   Measurement Instruments to Assess Treatment Acceptability ....................................... 35

      Treatment Acceptability Measures for Clinical Treatments ..................................... 36

         Treatment evaluation inventory ........................................................................ 36

         Treatment evaluation inventory-short form ....................................................... 36

         Treatment acceptability rating form ................................................................. 37

         Treatment acceptability rating form-revised ................................................... 38

      Treatment Acceptability Measures for School-Based Interventions ..................... 38

         Intervention rating profile .................................................................................... 39

         Intervention rating profile-15 ............................................................................. 40

         Abbreviated acceptability rating profile ............................................................... 40

         Children's intervention rating profile .................................................................. 41

         Behavior intervention rating profile ..................................................................... 41

         Assessment rating profile ..................................................................................... 42
## Assessment rating profile-revised ................................................................. 43
## Curriculum based measurement-acceptability scale ........................................... 43
## Acceptability of Reading Interventions ............................................................ 45
## Purpose of the Study ........................................................................................ 51
##
## II. METHOD ...................................................................................................... 53
## Overview ........................................................................................................... 53
## Research Questions .......................................................................................... 53
## Sampling Procedures ....................................................................................... 54
## Determination of Sample Size ......................................................................... 54
## Recruitment ...................................................................................................... 55
## Development of the RIRP ................................................................................ 57
## The Reading Intervention Rating Profile ......................................................... 58
## Item development of the RIRP ....................................................................... 58
## Demographic questions .................................................................................... 60
## Case scenario .................................................................................................... 60
## Validity .............................................................................................................. 62
## Expert review ................................................................................................... 62
## Data Collection ............................................................................................... 63
## Research Design .............................................................................................. 65
## Exploratory Factor Analysis Procedures ......................................................... 66
## Factor extraction ............................................................................................... 69
## Factor rotation .................................................................................................. 70
## Cronbach’s Coefficient Alpha ......................................................................... 71
## Independent Sample t-Test Procedures ........................................................... 72
## Mann-Whitney U Test Procedures ................................................................... 73
##
## III. RESULTS .................................................................................................... 76
## Overview ........................................................................................................... 76
## Descriptive Statistics ....................................................................................... 76
## Participants ....................................................................................................... 77
## Participants demographics ............................................................................... 77
## Professional training & experience - all participants ....................................... 78
## Instructional/organizational variables - all participants .................................... 79
## Special Education Teacher Subsample ............................................................. 80
## Special education teacher demographics ....................................................... 81
## Professional training & experience - special education teachers .................... 81
## Instructional/organizational variables - special education teachers ................. 82
## General Education Teacher Subsample ............................................................ 83
## General education teacher demographics ....................................................... 83
## Professional training & experience - general education teachers .................... 84
## Instructional/organizational variables - general education teachers ............... 85
## Results from Data Analyses ............................................................................. 86
Exploratory Factor Analyses Results........................................................................86
Test of statistical assumptions .................................................................................86
Factor extraction and rotation ..................................................................................87
Cronbach’s Coefficient Alpha Results .....................................................................90
Independent Sample t-Test Results ..........................................................................94
The Mann-Whitney U Test Results ..........................................................................95

IV. DISCUSSION .......................................................................................................100
Overview ..................................................................................................................100
Review of Statement of Problem and Study Purpose .............................................100
Discussion of the Study Findings .............................................................................102
Exploratory Factor Analyses and Reliability of the RIRP ........................................102
Relationships Between Teacher Characteristics and RIRP Ratings ......................104
  Years of teaching experience .................................................................................104
  Level of education ................................................................................................106
  Special education versus general education .........................................................108
Limitations ................................................................................................................111
  Psychometric Limitation ........................................................................................111
  Methodological Limitations ...................................................................................111
  Statistical Limitations ............................................................................................113
Implications for Future Research ............................................................................114
Conclusion ...............................................................................................................116

APPENDICES ..........................................................................................................119
APPENDIX A. LETTER OF CONSENT .................................................................120
APPENDIX B. READING INTERVENTION RATING PROFILE (RIRP) ..................123
APPENDIX C. DEMOGRAPHIC QUESTIONS .........................................................125
APPENDIX D. INTERVENTION CASE STUDY .....................................................129
APPENDIX E. EXPERT EVALUATION SHEET .....................................................132

REFERENCES ..........................................................................................................134
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Witt and Elliott’s (1985) Model of Treatment Acceptability</td>
<td>20</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Participant Demographics</td>
<td>78</td>
</tr>
<tr>
<td>2. Professional Training &amp; Experience, All Participants</td>
<td>79</td>
</tr>
<tr>
<td>3. Instructional/Organizational Variables, All Participants</td>
<td>80</td>
</tr>
<tr>
<td>4. Special Education Teacher Demographics</td>
<td>81</td>
</tr>
<tr>
<td>5. Professional Training &amp; Experience, Special Education Teachers</td>
<td>82</td>
</tr>
<tr>
<td>6. Instructional/Organizational Variables, Special Education Teachers</td>
<td>83</td>
</tr>
<tr>
<td>7. General Education Teacher Demographics</td>
<td>84</td>
</tr>
<tr>
<td>8. Professional Training &amp; Experience, General Education Teachers</td>
<td>85</td>
</tr>
<tr>
<td>9. Instructional/Organizational Variables, General Education Teachers</td>
<td>86</td>
</tr>
<tr>
<td>10. Factor Matrix from the RIRP for Special Education Teachers</td>
<td>88</td>
</tr>
<tr>
<td>11. Factor Matrix from the RIRP for General Education Teachers</td>
<td>90</td>
</tr>
<tr>
<td>12. Descriptive Statistics for the RIRP for Special Education Teachers</td>
<td>91</td>
</tr>
<tr>
<td>13. Descriptive Statistics for the RIRP for General Education Teachers</td>
<td>93</td>
</tr>
<tr>
<td>14. Mean Ranks and Sum of Ranks for the Ratings of General Education Teachers on the RIRP – Experience</td>
<td>96</td>
</tr>
<tr>
<td>15. Mean Ranks and Sum of Ranks for the Ratings of Special Education Teachers on the RIRP – Level of Education</td>
<td>97</td>
</tr>
<tr>
<td>16. Mean Ranks and Sum of Ranks for the Ratings of General Education Teachers on the RIRP – Level of Education</td>
<td>98</td>
</tr>
</tbody>
</table>
17. Mean Ranks and Sum of Ranks for the Ratings of Special Education Teachers and General Education Teachers on the *RIRP* ........................................ 99
CHAPTER 1
REVIEW OF THE LITERATURE

Introduction to the Study

In this chapter, I synthesize research and background literature to provide a rationale for my study. The purpose of my study was to develop and validate a rating scale to measure teachers’ treatment acceptability of reading interventions for elementary students struggling with reading, including students with learning (specifically, reading) disabilities. Also, I examined whether teachers’ ratings on the newly developed scale differ depending on teachers’ certain characteristics.

In this chapter, I first describe the prevalence, types, and impact of reading difficulties on academic development in elementary grades. Second, I overview research-based practices known to be effective in remediating reading deficits in elementary schools and discuss the importance of these practices. Third, I discuss issues related to the research-to-practice gap in education and special education, and factors that contribute to the research-to-practice gap. Fourth, I examine the concept of social validity in research and practice, with a specific focus on models of treatment acceptability. Fifth, I review research on the treatment acceptability of behavioral interventions, factors that impact treatment acceptability, and measures of treatment acceptability. Last, I discuss the treatment acceptability of reading interventions and measures of treatment acceptability for reading. Finally, I discuss the need for a reliable and valid measure of treatment acceptability of research-based reading interventions for struggling readers and
the need for examining whether teachers’ acceptability of interventions vary regarding teachers’ characteristics.

**Prevalence, Types, and Impact of Reading Difficulties in Elementary Grades**

Reading difficulties in elementary grades often lead to undesirable academic and developmental outcomes (Hitchcock, Prater, & Dowrick, 2004). Therefore, it is important to know the prevalence and types of reading problems in schools. In this section, I first describe the prevalence of learning disabilities involving language and reading. Next, I discuss the nature of reading problems and reasons why some students struggle to learn to read. Finally, I synthesize research examining the impact of reading difficulties on academic and other developmental outcomes.

**Prevalence and types of reading problems.** In 1963, Samuel Kirk created the term *learning disabilities* in order to clarify the special characteristics of this group (Heward, 2009). Over the years, the number of students identified as having learning disabilities has grown rapidly and has been one of the fastest growing categories of special education. On the other hand, while the identification rate in the learning disability category increased more than 300% between 1976 and 2000, it has dropped each year since 2002 by nearly 2% (Cortiella & Horowitz, 2014). Currently, this category represents almost 42% of all students served in special education in the United States (The Individuals with Disabilities Act [IDEA], 2011).

Lyon (1995) estimated that 80% of students with learning disabilities have problems in reading. The IDEA (2004) defines *reading disability* as significant underachievement in reading that is not attributable to other disabilities or poor
instruction. Students with reading disabilities have poor reading skills in spite of having average to above average intelligence (Snowling, 2000). Such students may have trouble with vocabulary, decoding, spelling, comprehension, and fluency. The National Reading Panel Report (NRP, National Institute of Child Health and Human Development [NICHHD], 2000) identified five essential areas of focus for effective reading instruction; phonemic awareness, phonics, fluency, vocabulary, and comprehension.

A phoneme refers to the smallest unit in the spoken word. Phonemic awareness refers to awareness of the individual sounds in spoken words (NICHHD, 2000) and being able to manipulate these sounds. For example, *cat* has three sounds; recognizing that “cat” is made up of three sounds (/k/, /a/, & /t/) is phonemic awareness. Being able to manipulate these component sounds is an important aspect of phonemic awareness. Manipulating phonemes includes separating sounds in a word, blending them, and taking them apart again (NICHHD, 2000).

Phonics refers to the relationship between a letter in a specific word and its sound (Chard & Osborn, 1999). That is, letters are associated with specific sounds. It is important to note that phonics is related to written language and has a set of rules or principles. Students who understand the principles of phonics might know how to read an unknown word when they see it (Chard& Osborn, 1999; NICHHD, 2000). Students who have phonological awareness are also good at using rhymes and decoding text (Chard & Dickson, 1999).

*Fluency* is another critical area in reading. The NRP report defines *fluency* as the ability to read accurately, quickly, and with correct expression (NICHHD, 2000).
Students who are not fluent tend to have trouble reading comprehension (Kim, Wagner, & Foster, 2011). Reading fluency is important because a fluent reader can focus on the meaning of what is being read without distraction and thus can comprehend the information readily (Hudson, Lane, & Pullen, 2005).

*Vocabulary* is also an important element in reading. Children increase their vocabulary by listening, speaking, writing, and reading. Moreover, direct instruction or indirect exposure to vocabulary can increase the number of words that children know (Coyne, Zipoli, & Ruby, 2006). Through early instruction and intervention, children can enhance their vocabulary knowledge (Cuticelli, Coyne, Ware, Oldham, & Rattan, 2014).

Finally, *comprehension* can be simply defined *understanding the meaning* of what is being read. Comprehension emerges from the interaction between the reader and the text (NICHHD, 2000). Comprehension refers to the process of learning or gaining new information from what is read. Students who have good comprehension skills focus on the meaning rather than the structure of the text. In other words, comprehension involves reading to learn rather than learning to read. Individuals who are good at comprehension can identify the main idea of a passage, ask questions, and answer when a question is presented to them (Coyne et al., 2006).

All five reading skills are essential for good reading ability (Cummings, Kaminski, Good, & O'Neil, 2010). Being a good reader requires effective and efficient use of all five of these reading elements interactively (McCardle, Scarborough, & Catts, 2001) since all these reading skills are related to each other. For example, fluency and comprehension are two important interrelated concerns for students who are experiencing
difficulties in reading (Hudson et al., 2005; Therrien & Hughes, 2008); indeed, fluency is a strong predictor of students’ difficulties in reading comprehension (Stanovich, 1991). Similarly, phonological awareness contributes to reading fluency, as it requires understanding the relationship between phonics and letters, thereby facilitating rapid identification of words (NICHHD, 2000). Additionally, good vocabulary instruction increases reading comprehension (Kameenui, Carnine, & Freschi, 1982), while vocabulary knowledge enhances fluency (Nation, 1993). Metsala (1999) has indicated that vocabulary growth is associated with both phonological awareness and non-word repetition while Scarborough (1998) found that phonological awareness and expressive vocabulary are important predictors of later reading skills. Roberts (2005) also found that the size of one’s vocabulary is a significant contributor to phonemic awareness. Clearly, all the components of reading are interrelated and they affect students’ overall reading ability that impacts students’ educational and developmental outcomes.

**Impact of reading on educational and developmental outcomes.** A significant number of students experience reading problems in elementary grades and lag behind their peers in academic achievement. According to a report by the National Center for Education Statistics (NCES, 2004), the reading achievement of 37% of fourth-grade students was below grade level. Moreover, Francis, Shaywitz, Stuebing, Shaywitz, and Fletcher (1996) stated that approximately 75% students who demonstrated difficulties in reading in third grade continued to read poorly until the end of high school.

Difficulties in reading negatively affect achievement in math and other academic subject areas. Research indicates that there is a positive relationship between
phonological awareness and mathematical computational skills (Hecht, Torgesen, Wagner, & Rashotte, 2001). Difficulties in reading comprehension might affect understanding math problems (Fuentes, 1998) as well as other content areas (Billmeyer & Barton, 1998). As a result of these difficulties, students with reading disabilities have lower grades and higher course failure rates than those without reading disabilities; thus, students with reading disabilities have high risk of grade retention and school dropout (Cortiella & Horowitz, 2014).

In addition to experiencing problems in academic areas, students with reading disabilities are at increased risk problems in social and emotional development. Students with reading disabilities may struggle in social interaction and have low self-esteem due to their academic hardships (Friend, 2008). Because of these struggles, they may have problems making friends, leading them to feel lonely and depressed well into adulthood (Maag & Reid, 2006). Also, behavior problems might result from these negative experiences and academic failure (Farmer, 2000).

The pervasive and long-lasting negative impact of reading problems in early grades highlights the importance of providing timely, effective early reading interventions in schools. Prevention, early intervention, and remediation are necessary for students with reading disabilities in order to help them succeed both academically and developmentally (Lyon, 2003). If reading problems at an early age are not prevented or remediated, failure in reading as well as in other content areas is very likely. As a result, underachievement and dropout rates increase (Denti & Guerin, 1999). Therefore, before students with reading problems or at risk for reading failure face long-term negative
consequences, educators need to be aware of the importance of prevention, early intervention, and remediation. Unfortunately, commonly used practices to teach reading often do not improve students’ reading outcomes. Many studies have shown that many interventions offered to students with or at-risk for reading disabilities are not effective enough to increase their reading skills (Schumaker, Deshler, & Ellis, 1986; Torgesen, 2004; Torgesen et al., 2001). To reading and related outcomes for students with reading disabilities, interventions need to remediate their skills by being more intensive, explicit, and systematic.

**Effective Reading Interventions in Special Education**

In this section, some selected interventions for students with reading problems will be described, including those related to phonological awareness, phonemic awareness, fluency, vocabulary, and comprehension. This overview includes effective interventions for students with and at-risk for reading disabilities.

Providing effective interventions to students with reading problems is a necessary step to increase their success. Although many teachers want to implement effective practices to increase their students’ reading outcomes, identifying effective interventions may be complicated by uncertainty regarding the term “best practices.” Kauffman (1996) stated that many teachers have actually been using practices that are ineffective due to ambiguity regarding the definition of an “effective” practice. While some effective practices exist, others lack either sufficient quality or quantity of empirical evidence (Greenwood & Abbott, 2001). Increasing the quality, usability, and accessibility of research will prevent and remediate the deficits in reading.
In the last decade, the report of NRP has identified effective practices for students with reading problems (NICHHD, 2000). According to the NRP report, instruction in phonemic awareness is necessary for beginning reading. Such instruction focuses on phoneme blending and segmenting, which are the important skills for early grades. Many studies have indicated that phonemic awareness instruction is not only effective but also better than methods that deemphasize directly teaching phonemic awareness, e.g., whole-language instruction (Bond & Dykstra, 1967; NICHHD, 2000; Snow, Burns, & Griffin, 1998). Effective beginning reading instruction includes explicit and systematic procedures that help students clearly to understand alphabetic principles. The NRP determined that phonemic awareness instruction maximizes outcomes if it is implemented with kindergarten students at least six minutes per day, or thirty minutes per week, for a total time of 18 hours (NICHHD, 2000).

Another reading component is phonological awareness. The NRP has identified systematic and explicit phonics instruction as an important element of effective reading instruction (see also Foorman & Torgesen, 2001). Phonics instruction includes two components: synthetic phonics and larger-unit phonics. Synthetic phonics, which features matching and blending, involves matching sounds to letters and integrating the sounds to make words. Larger-unit phonics involves onsets and rimes that are bigger than phonemes. In this case, students divide words into letter chunks, decode them, and blend that word pieces (Ehri, Nunes, Stahl, & Willows, 2001). According to the NRP, systematic phonics instruction produces positive outcomes for students in kindergarten and/or first grade and provides substantial improvement in reading through the sixth
grade (NICHHD, 2000). For example, O'Connor, Jenkins, Leicester, and Slocum (1993) investigated the effects of teaching phonological awareness skills to preschool students with disabilities. Students were given phonological tasks including rhyming, blending, and segmenting over a period of seven weeks, four times a week for 10 minutes each day. The results showed that the students who were in the experimental group had learned significantly more than students in the control group.

In addition to phonemic awareness and phonological awareness, fluency instruction is essential for teaching students to read. The report of the NRP considers fluency to be a vital part of reading skills, with repeated practice viewed as a significant way to increase fluency (NICHHD, 2000). One of the recommended practices for fluency is repeated reading, which involves a series of repeated readings by students (Samuels, 1979). The aim of repeated reading is to increase the number of words that a student can read per minute (Samuels, 1979). The intervention requires students to reread a given passage until they reach the criterion of fluency for that particular occasion (Therrien, 2004). In all such activities, active student participation in the instruction is one of the important requirements for enhancing student performance and achievement. A meta-analysis of reading intervention research indicated that students with and without learning disabilities can benefit from repeated reading intervention (Yang, 2001) showing that repeated reading is an instructional approach that increases students’ reading fluency (Alber-Morgan, 2006).

Vocabulary is another essential element for reading. The NRP has provided various vocabulary teaching practices that are useful for students with reading problems,
such as explicit vocabulary instruction and reading aloud (NICHHD, 2000). For effective vocabulary instruction, word definitions should be explicit, examples and non-examples should be carefully chosen, and indirect word learning should be promoted by teaching students how to use the dictionary, how to determine word roots, and how to interpret context clues (Sedita, 2005). Reading aloud is another practice that helps students learn unfamiliar words. This practice includes talking about what will be read before reading aloud, during reading aloud, and after reading aloud, and discussing any new words in the story. In this way, students can improve their vocabulary levels (Santoro, Chard, Howard, & Baker, 2008).

The final component of reading instruction is comprehension. There are many useful comprehension practices such as answering questions, monitoring comprehension, using graphic organizers, listening actively, and calling on prior knowledge (Trabasso & Bouchard, 2002). Prompting students to identify and answer the five “WH” questions (who, what, where, when, and why) is one of the reading practices for reading comprehension (Clark, Deshler, Schumaker, Alley, & Warner, 1984). Each “WH” question has a symbol. Students read the given passage by asking “WH” questions and then answering them. Finally, they mark their answers with the appropriate symbol. An additional way of aiding comprehension is to provide students with well-structured small group discussions, where students can talk about their personal experiences and activate their background knowledge by listening to what others say. Clearly, this is another way that teachers can increase students’ comprehension levels and get positive reading outcomes (Goldenberg, 1993).
As research suggests, well-supported interventions are important both to prevent and to remediate reading problems. In particular, students who have reading disabilities must receive high quality interventions in order to maximize their potential. Well-supported reading interventions are not only beneficial for students with reading disabilities, but they are also critical for minority students and those from low socioeconomic backgrounds, groups that are observed to be at higher risk for reading problems (Torgesen, 2004). These students are more likely to need evidence-based practices in reading that are explicit, direct, and systematic than their peers (Torgesen, 2004).

The report of NRP has determined effective practices for students with reading problems (NICHHD, 2000) and Denton, Vaughn, and Fletcher (2003) provided a brief summary of the report results and stated “…the following areas must be addressed in order to enhance the reading development of all students; (1) an effective and knowledgeable teacher, (2) integration of key instructional components, (3) differentiated instruction for students with reading difficulties, (4) explicitness of instruction, and (5) bridging the gap between research and practice” (p. 202). Evidence based practices are seen as the bridge between research and practice (Cook, Tankersley, & Landrum, 2009). It is necessary, however, to look factors that contribute to the research to practice gap in order to identify and eliminate barriers to the implementation of well-supported interventions.
Research to Practice Gap

Research has helped practitioners achieve desired outcomes in such fields as medicine, engineering, and agriculture (Jones, 2009) but in education, the use of empirically validated effective practices is less common (Kauffman, 1996). The report of the National Research Council (NRC, 2002) emphasized using scientific methods in education similar to other fields or professions (such as geophysics and political science). In special education, the gap between research and practice is also an ongoing concern (Cook & Odom, 2013). While the focus of special education is using research based practices and applying direct and frequent assessments to determine the impact of instruction (Jones, 2009), numerous studies indicate that students with disabilities receive ineffective instruction (Blackorby & Wagner, 1996; Jones, 2009). Spencer, Detrich, and Slocum (2012) summarized the research to practice gap by stating that “consumers [students] are not receiving services that are based on the best research evidence that exists and therefore may suffer from poorer outcomes and unnecessary costs associated with ineffective treatments” (p. 128).

More and more, schools are implementing a Response to Intervention (RtI) model as a systematic way to implement and sustain research-based instruction for struggling students. RtI uses multiple “tiers” to provide a series of high-quality interventions in reading and math for struggling learners (Fuchs & Fuchs, 2006). Thus, RtI is a multi-tiered system focused on early identification and prevention strategies for struggling learners, designed to provide an effective support to students before they fail (What Works Clearinghouse [WWC], 2009). In each tier, students receive research-based
instruction; the intensity of intervention at each tier increases to match students’ needs. Fuchs and Fuchs (2006) stated that the reading and math instruction can be made more intensive by “(a) using more teacher-centered, systematic, and explicit (e.g., scripted) instruction; (b) conducting it more frequently; (c) adding to its duration; (d) creating smaller and more homogenous student groupings; or (e) relying on instructors with greater expertise” (p. 41).

While RtI has many strengths (i.e., providing immediate instructional support, early intervention, and high-quality instruction), there are also limitations that may contribute to the research to practice gap. For example, general education teachers’ background knowledge and skills may not be sufficient for implementing the scientifically-based instruction required in an RtI model (Mastropieri & Scruggs, 2005). Also, there is uncertainty regarding how implementation fidelity is determined in RtI procedures (Gerber, 2005). These limitations indicate that there are barriers that affect the use of these scientific methods in education. A study by Vanderlinde and van Braak (2010) identified four important barriers affecting the use of research by practitioners: inadequate *applicability* of the research; *vagueness* of the research material; *complexity* and *technical language* of the research; and the *nature* of the research itself.

Regarding inadequate applicability of research, some researchers have stated that research-based practices can be difficult to apply in educational settings (Broekkamp & van Hout-Wolters, 2007; Burkhardt & Schoenfeld, 2003). Practical concerns are often not adequately addressed in educational research. Furthermore, research is frequently conducted in an environment that is different from real school settings, limiting the
generalizability of research results to school settings (Greenwood & Abbott, 2001). Carnine (1997) discusses the quality of the research and states that research needs to be *useable*; that is, a study should be practical and clearly written so as to help practitioners follow its steps and easily apply the results in practice. However, the report of the Coalition for Evidence-Based Policy (2002) states that educational research is often of poor quality and exerts little influence on educational practice. Therefore, many practitioners believe that educational research offers little practical guidance and that research-validat ed interventions do not produce outcomes that practitioners value (Deshler, 2003).

Another contributor to the research-practice gap is that educational research does not consistently yield valid and reliable findings (Broekkamp & van Hout-Wolters, 2007). Carnine (1997) also discusses that a study should be *trustworthy*, meaning that educators can have confidence in the research findings. A poor track record of trustworthiness contributes to practitioner wariness toward educational research. For example, the study of Boardman and his colleagues (2005) indicated that many teachers feel skeptical in terms of the research validity. Also, teachers reported that the evidence of effectiveness of recommended instructional practices is not provided to them (Boardman, Argüelles, Vaughn, Hughes, & Klingner, 2005) and research results are often unclear and even contradictory (Vanderlinde & van Braak, 2010).

The technical and obscure language of research presents yet another barrier for teachers. Results of a study by Vanderlinde and van Braak (2010) indicated that teachers found the technical and complex language of research reports to be unintelligible. While
practitioners ask for clear and concise information, researchers use very complicated style in writing about their research (Mortimore, 2000). In the same study (Mortimore, 2000), researchers stated that using domain-specific jargon is easier than writing in teacher-friendly terms. They also added that their work is generally judged in relation to publication in high-impact professional journals (where domain-specific technical language is valued), not in relation to dissemination through teacher-friendly outlets (Jochems, 2006).

Last, the nature of research is seen as another important barrier. For example, descriptive studies (common in educational research) do not provide guidance for teachers regarding classroom practice (for example, what teachers should do to enhance their teaching). Teachers generally need and value practices that are demonstrably effective (Vanderlinde & van Braak, 2010). Also, special education teachers reported that the research studies generally provide information for general education students and there is insufficient research focusing on students with disabilities (Boardman et al., 2005).

In addition to these barriers, Greenwood and Abbott (2001) stated that practitioners have limited opportunities to improve their professional development. The general disconnect between researchers and practitioners leads practitioners to see research as irrelevant to their personal experiences in the classroom. The reason of this disconnect may be that researchers do not sufficiently partner with practitioners in the research process (Greenwood & Abbott, 2001). Also, practitioners are often not
sufficiently trained to implement the results of an educational research study (Broekkamp & van Hout-Wolters, 2007).

Carnine (1997) states that accessibility is another indicator of research quality. It is known that many practitioners have difficulty obtaining the necessary information about a particular practice and interpreting the research findings (Carnine, 1997). Therefore, research studies need to be accessible if practitioners are to obtain the necessary information. All these factors contribute to the failure to bridge the research-practice gap that is associated with compromised implementation of effective practices, and consequent poor student outcomes (Cook, Smith, & Tankersley, 2012).

The gap between research and educational practice is a complex phenomenon (Vanderlinde & van Braak, 2010). While many factors contribute to the research to practice gap, social validity is particularly important. Teachers’ beliefs about the value or worth of a treatment can affect their use of the treatment in the classroom (Chafouleas, Briesch, Riley-Tillman, & McCoach, 2009) and, as such, is directly related to the research to practice gap.

Social Validity

Social validity, or the user’s subjective evaluation of an intervention’s value or worth, is an important issue when considering high quality educational interventions, because the user’s evaluation determines the level of implementation (Reimers, Wacker, Cooper, & DeRaad, 1992). In 1978, Wolf discussed the importance of social validity in regards to behavioral treatments, and stated that the treatments needed to be valued by practitioners to bring practitioners into the science. In his article, Wolf emphasized the
necessity of subjective evaluations by gathering data that measures social validity, judgments of consumers (Foster & Mash, 1999).

Social validity is defined as judgments by intervention consumers of the importance and acceptability of three main aspects of a treatment (or intervention): treatment goals, treatment effects, and treatment procedures (Kazdin, 1977; Wolf, 1978). The social validity of treatment or intervention goals reflects consumer/societal approval or disapproval of the purpose or objectives of intervention. The social validity of treatment goals can be assessed at various levels: “broad social goals, behavioral categories, and discrete responses within categories of behavior” (Fawcett, 1991, p. 235). Assessing broad treatment/intervention goals involves surveying or otherwise measuring the thoughts and opinions of consumers who make the decision as to whether an outcome is significant (Fawcett, Seekins, & Jason, 1987). While Fawcett and his colleagues (1987) focused their discussion on the social significance of behavioral goals, their conceptualization of social validity can also be applied to the broad or specific goals of academic instruction and/or interventions.

The social significance of treatment effects is based upon consumers’ satisfaction with actual outcomes, including planned and unplanned outcomes (Wolf, 1978). Similar to goals, the social significance of treatment or intervention effects can also be assessed in relation to various levels: proximal effects, intermediate effects, and distal effects (Fawcett, 1991). Proximal effects include the social significance of the main effect of the intervention (e.g., increased rate of reading fluency). Intermediate effects are indirect effects that are influenced by the increase in the proximal effect (e.g., increased rate of
reading comprehension related to reading fluency). Distal effects reflect outcomes such as increased academic achievement. The assessment of social validity must include all levels of treatment effects in order to specify their acceptability by consumers.

The social validity of treatment procedures (commonly referred to as treatment acceptability) is the third major component of social validity (Wolf, 1978). The social validity of treatment or intervention procedures can also be assessed at various levels: “comprehensive treatments programs, treatment packages or components of a program, and specific behavioral procedures or components of treatment packages” (Fawcett, 1991, p.236). Wolf (1978) defined procedural social validity as the acceptability of treatment procedures to consumers. Wolf (1978) discussed the relation between social validity and teachers’ acceptance of an intervention and indicated that treatment acceptability is one of the significant elements of social validity; therefore, it has a critical role in implementing educational practices.

**Treatment Acceptability**

In 1980, Kazdin acknowledged treatment acceptability as a key component of social validity and stated that research on social validity is primarily about the acceptability of treatment procedures. Researchers in applied behavior analysis use the term treatment acceptability to refer to the acceptability of the use of certain behavior modification strategies or specific behavior change interventions, or even to particular classroom interventions (Polloway, Bursuck, Jayanthi, Epstein, & Nelson, 1996). Kazdin (1981) defines treatment acceptability as “judgments by laypersons, clients, and others of whether treatment procedures are appropriate, fair, and reasonable for the problem or
client” (p. 493). In other words, treatment acceptability refers to consumer perception of usefulness, fairness, and practicality of the intervention in a specific setting. Other researchers besides Kazdin have developed conceptual models to better identify and understand issues that affect treatment acceptability. Conceptual models show possible relationships among influencing factors and try to explain how treatment acceptability is related to these factors. In the next section, I provide an overview of these conceptual models of treatment acceptability.

**Conceptual Models of Treatment Acceptability**

Several researchers have delineated the critical dimensions of treatment acceptability and developed conceptual models to indicate the relationships between these dimensions. In this section, the conceptual models of treatment acceptability are described in order to demonstrate multiple aspects of treatment acceptability and factors related to treatment acceptability.

**Witt and Elliott’s (1985) model.** Witt and Elliott’s (1985) conceptual model of treatment acceptability is based on four critical elements: “acceptability, use, integrity, and effectiveness” (p. 274). In this model, the relationship among these elements is sequential and reciprocal. Acceptability is considered the initial element in the sequence because it affects treatment use and integrity. After a treatment is perceived as acceptable, it is expected that its use will be high. The link between use and effectiveness is treatment fidelity. When a treatment is implemented with fidelity, its effectiveness will also be high (Witt & Elliott, 1985). Furthermore, if there is a break in any of the connections, the remaining connections are also affected (See Figure 1). Some studies
provide empirical support for this model (Besalel-Azrin, Azrin, & Armstrong, 1977; Kazdin, 1981; Kirigin, Braukmann, Atwater, & Wolf, 1982).

Reimers, Wacker, and Koeppl’s (1987) model. Reimers, Wacker, and Koeppl (1987) provided a more complex conceptual model of acceptability. Reimers and his colleagues (1987) hypothesized that understanding of the treatment is the initial component of the model since the treatment acceptability is a concern when consumers know what the treatment is. Treatment acceptability must be assessed in relation to whether the treatment or intervention is well understood by consumers. After they have knowledge about and understanding of the treatment, other components in the model come into play. The model describes the hypothetical relationship among various components such as understanding, compliance, acceptability, effectiveness, and maintenance.

Reimers and his colleagues (1987) explain the model as follows. When poor understanding of a treatment occurs, it may lead to low adherence to the treatment steps which reduces the probability of treatment effectiveness. Therefore, there is a need to
provide information about the recommended treatment so consumers understand it better. On the other hand, good understanding might be associated with either low acceptability or high acceptability, which can also affect adherence to treatment. While low acceptability is associated with low adherence to intervention steps (and low effectiveness as a result), high acceptability is associated with high adherence. In turn, high adherence might be followed by low effectiveness or high effectiveness and might influence the components of maintenance and family or classroom disruption (Reimers et al., 1987). Once the recommended treatment has been tried, its continued use (maintenance) is the major concern. At the end, there may be a need to modify the treatment or provide a new treatment to the consumers (See Figure 2).

Figure 2. Reimers, Wacker, and Koeppl’s (1987) Model of Treatment Acceptability
**Elliott’s (1988) model.** Elliott (1988) stated that neither Witt and Elliott’s model (1985) nor Reimers and his colleagues’ model (1987) adequately explain the complex relationship among relevant variables. Building on Reimer’s et al. (1986), Elliott (1988) proposed a possible relationship between low acceptability and high adherence or high acceptability and low adherence. Elliott only made a small modification to the model of Reimers and his colleagues, and indicated that low acceptability might be followed by high adherence and high acceptability might be followed by low adherence. Other than showing these possible relationships, he did not make any changes to the model of Reimers et al. (1987) (See Figure 3).

On the other hand, Elliott, Witt, Kratochwill, and Stoiber (2002) suggested four factors that affect the use and selection of interventions, namely: *intervention effectiveness, intervention acceptability, intervention resistance and empowerment,* and *intervention integrity.* First, an intervention needs to be effective in order to be selected. Elliott and his colleagues (2002) stated that the first question usually asked by researchers or practitioners is whether an intervention works or not. If the intervention does not sufficiently effective, it could mean that the implementation of the intervention has failed (Baer, Wolf, & Risley, 1968). The practical impact of the intervention affects its acceptability; thus, the intervention needs to produce effective results and practical value to be deemed socially valid.

However, Elliott et al. (2002) stated *effectiveness of an intervention* alone is not sufficient to recommend its use in practical settings; thus, its procedures have to be *acceptable* by consumers. The intervention acceptance is one of the most important
factors to use and select an intervention; therefore, it is viewed as a central component of social validity and it has obtained significant interest among the professionals such as researchers and educators (Elliott et al., 2002). The selection of interventions by teachers and the implementation of interventions in the classroom mostly depend on their acceptability. Other than teachers’ judgments of treatment acceptability, legal and ethical entities such as institutional review committees also evaluate particular interventions to determine whether the procedures involved are acceptable or not (Elliott et al., 2002).

Also, intervention resistance is an important construct for understanding teachers’ acceptance of particular interventions. Sometimes teachers never even try out the intervention or reject it after implementing it only briefly (Elliott et al., 2002). Teachers must be empowered and supported in their use of interventions to counteract this resistance. Support must be provided by consultants and school psychologists who can assist teachers in terms of determining teacher needs and providing resources to meet their needs. On the other hand, intervention resistance might occur when intervention packages are installed without taking into account teacher skills and classroom resources (Elliott et al., 2002), leading them to reject the suggested practice and criticize the procedures of the intervention.

Last, intervention integrity affects intervention selection and use. Elliott et al. (2002) defined intervention integrity as “the degree to which an intervention plan is implemented as intended” (p. 247). Interventions are highly acceptable to teachers and applied with integrity when the interventions (a) are easy to use, (b) require less time, (c) are positive, (d) are deemed effective, and (e) are well matched with teacher skills and
available resources (Telzrow & Beebe, 2002). When interventions require multiple resources and materials, the integrity of interventions can be difficult to establish and their use might not be acceptable to teachers (Perepletchikova & Kazdin, 2005). Conversely, the unacceptability of a treatment can lead to low intervention integrity.

Clearly, intervention effectiveness, intervention acceptability, intervention resistance, and intervention integrity are critical; however, these factors are infrequently considered in efforts to improve outcomes of classroom interventions (Elliott et al., 2002). To achieve successful intervention outcomes, all of these factors and their interrelationships must be considered.

*Figure 3.* Elliott’s (1988) Model of Treatment Acceptability
In summary, conceptual models of treatment acceptability provide various ways to understand what treatment acceptability is and the factors that influence treatment acceptability. To guide the intervention application process, all the components need to be considered. However, research indicates that there are many factors that affect treatment acceptability in addition to the components indicated in these conceptual models. In the next section, these moderating and mediating factors are summarized. Also, I provide an overview of instruments that have been used to measure treatment acceptability and the factors that influence treatment acceptability.

**Synthesis of Research on Treatment Acceptability**

While the focus of my study is on the treatment acceptability of reading interventions, most of the existing research has focused on the treatment acceptability of behavioral interventions. In this section, I summarize and synthesize the research on treatment acceptability of behavioral interventions. First, I summarize research on factors that influence treatment acceptability. Next, I overview instruments used to measure treatment acceptability. Finally, I review studies that investigate the treatment acceptability of reading interventions for students with reading difficulties.

I conducted a systematic literature review to uncover research on treatment acceptability in general education and special education across different types of practices. I conducted electronic searches of educational databases such as ERIC, EBSCO, JSTOR, and Psych INFO, along with Google Scholar. I searched the literature using different key word combinations such as “treatment acceptability,” “intervention
acceptability,” “teacher acceptability,” “teacher perspectives,” “teacher attitudes,” and “treatment acceptability measurement instruments.”

**Factors Influencing Treatment Acceptability**

For clinical (i.e., non-school) populations, Kazdin and his colleagues identified many factors that affect treatment acceptability, including treatment effectiveness (Kazdin, 1981), adverse side effects of an intervention, and the use of technical jargon in research studies (Kazdin & Cole, 1981). In educational settings, treatment acceptability is just as important as in clinical settings. Research studies related to the acceptability of school-based interventions indicate that teachers are not passive recipients of recommended practices and may indeed reject the recommendations (Elliott, 1988; Witt & Martens, 1983). If an intervention is not perceived as acceptable, it is not likely to be applied in the classroom (Witt & Elliott, 1985).

Selected literature findings indicate that there are many factors that affect treatment acceptability. These factors can be categorized under four headings: **student factors, teacher factors, intervention factors, and contextual factors.** This section provides an overview of these factors.

**Student factors.** Studies show that child characteristics and the severity of behavior problems are the important factors influencing treatment acceptability of behavioral interventions. Martens and Meller (1989) found that child characteristics (intelligence, popularity) affect teachers’ ratings of the treatment acceptability of behavioral interventions. For example, teachers found it more acceptable to implement response-cost and home-based reinforcement procedures with students of average
intelligence than with students of low intelligence (Martens & Meller, 1989). Also, many studies showed that the severity of students’ problem behavior strongly influences teachers’ ratings of treatment acceptability. Consistent with the study of Kazdin (1980), Martens, Witt, Elliott, and Darveaux (1985) observed that as the severity of problem behavior increases, teachers’ treatment acceptability increases as well. Tingstrom (1990) obtained similar results, indicating that treatment acceptability was higher when time-out procedures were implemented with students presenting with severe problem behavior. Moreover, teachers rated low level of teacher involvement as less acceptable when students presented severe versus more mild behavior problems (Witt, Martens, & Elliott, 1984). Additionally, Elliott, Witt, Galvin, and Moe (1986) found that children rated behavioral interventions as more acceptable when the children targeted by such interventions exhibited more versus less severe behavior problems.

**Teacher factors.** Teachers’ certification type (general education, special education) and years of teaching experience also affect treatment acceptability. Maccini and Gagnon (2006) showed that, overall, special education teachers are more likely to approve of and implement recommended instructional strategies and instructional adaptations than general education teachers. Another study (Easton & Erchul, 2011) also found that special education teachers were more likely than general education teachers to accept and use monitoring and feedback methods as part of a behavior intervention plan. These findings suggest that special education teachers’ treatment acceptability are higher for some aspects of behavioral intervention as compared to general education teachers.

Years of teaching experience also affects the acceptability and implementation of
interventions in the classroom. Witt, Moe, Gutkin, and Andrews (1984) found that teachers who were highly experienced were less likely than less experienced teachers to accept recommended treatments. Witt and Robbins (1985) similarly found a negative relationship between years of teaching experience and acceptability of interventions. The results of the study by Power, Hess, and Bennett (1995) contradicted previous findings, indicating that knowledge and years of teaching experience were not connected to treatment acceptability ratings.

Professional development is another critical issue in treatment acceptability. For example, the study by Stormont, Reinke, and Herman (2011) investigated teachers’ knowledge about evidence-based practices with two groups of teachers—those with a graduate degree and those with an undergraduate degree. The results of the study showed that the group without a graduate degree rated non-evidence-based practices as evidence-based more often than those with a graduate degree. Moreover, according to Kutash, Duchnowski, and Lynn (2009), the absence of effective training and the inability to implement interventions with fidelity are the most significant barriers to implementing evidence-based practices. These findings suggest that professional development influences the acceptability and use of evidence-based practices in the classroom.

Additionally, the treatment acceptability of school-based interventions can vary based on who is expected to implement these interventions. Martens and colleagues (1985) investigated variables that affect teachers’ judgments of the acceptability of interventions. Two implementers of interventions (principal and teacher) were the variables rated by teachers. The findings of the study showed that teacher-directed
Interventions were more acceptable than principal-directed interventions, a conclusion supported by the earlier study by Witt and Robbins (1985). However, Tingstrom (1990) found that teachers rated teacher-directed interventions as less acceptable than interventions directed by school psychologists. Overall, it appears that who factors (who receives and implements the intervention) can influence judgments of the acceptability of behavioral interventions (Tingstrom, 1990). Next, I overview research examining what factors – what the treatment or intervention is or what it entails.

**Intervention factors.** In the studies that my search criteria, the type of treatment is the most influential variable affecting teachers’ acceptability of interventions. In these studies, researchers generally focused on behavioral interventions characterized as positive, reductive (Elliott, 1988), or pharmacological. Positive treatments generally included praise, daily behavior report cards, token economies, home-based reinforcement, and classroom lotteries. Reductive treatments included response-cost, time-out, ignoring, and overcorrection. Pharmacological interventions included stimulant medication for attention deficit hyperactivity disorder (ADHD).

In general, acceptability ratings for positive treatments were higher than those for negative treatment procedures. For instance, Elliott, Witt, Galvin, and Peterson (1984) indicated that teacher treatment acceptability ratings of praise, home-based reinforcement, and a token economy were higher than they were for response-cost lottery, time-out, and ignoring. These results have been replicated in other studies (e.g. Kazdin, 1980; Turan, Erbas, Ozkan, & Kurkuoglu, 2009; Von Brock & Elliott, 1987), with the researchers finding that reinforcement-based behavioral interventions were more
acceptable than reductive or punishment-based strategies. Similarly, Chafouleas, Riley-Tillman, and Sassu (2006) concluded that daily behavior report cards were deemed acceptable and useful by teachers. In sum, it can be said that teachers were more likely to accept and implement positive treatment procedures than negative ones. Kazdin (1981) also stated that the type of treatment influenced acceptability ratings, finding that, while the least acceptable treatment was medication, reinforcement-based strategies were the most acceptable approach. Participants’ acceptability rating order was positive practice first, time out second, and medication last. However, the generalizability of the study results is limited, because the study participants were mostly undergraduate psychology students (Elliott, 1988).

Power and colleagues (1995), obtained similar results, stating that daily report cards were significantly more acceptable than response cost or medication. On the other hand, while teachers were more likely to support positive practices, children’s ratings of behavioral interventions for their peers varied. According to Elliott et al. (1986), private teacher-student interactions, group reinforcement, and negative restrictions were the most acceptable interventions according to children’s ratings of treatment acceptability. Public admonition and negative group contingencies were perceived as unacceptable interventions.

The time required to implement an intervention is also a significant issue for teachers. The research on treatment acceptability suggests that the amount of time required to implement an intervention influences acceptability ratings by teachers (Martens et al., 1985; Witt et al., 1984). *Time efficiency* is a factor that affects teachers’
decisions to accept and use a behavioral intervention. de Mesquita and Zollman (1995) indicated that teachers prefer cognitive and cooperative-peer approaches, because they are time-efficient. Moreover, Elliott and his colleagues (1984) found that less time-consuming approaches are preferable to complex and more time-consuming interventions. Thus, teachers consider time as an important element in their decision to accept interventions and implement them in the classroom.

The theoretical orientation of a treatment is another factor that influences teachers’ decisions to use or not use an intervention. Witt (1986) has stated that teachers are more likely to choose the intervention if it is labeled as “humanistic education” rather than as “behavior modification.” Also, technical jargon tends to reduce treatment acceptability. Kazdin and Cole (1981) found that teachers preferred simple terms in research studies rather than highly technical jargon. Using complex language to explain an intervention negatively influenced participant ratings of the acceptability of an intervention (Kazdin & Cole, 1981).

The actual effectiveness of an intervention must also be considered as a potential influence on treatment acceptability. While some researchers found a relationship between the effectiveness of an intervention and its acceptability to teachers (e.g., Besalel-Azrin et al., 1977; Kirigin et al., 1982), Kazdin (1981) found no such relationship.

The jargon used to describe interventions can influence teachers’ ratings of treatment acceptability. For example, Witt et al. (1984) investigated the influence of the type of language used to report an intervention on teachers’ acceptance of that
intervention. These researchers used three sorts of jargon – pragmatic, behavioral, and humanistic. The findings of the study showed that a pragmatic description of an intervention was rated as more acceptable than a humanistic or behavioral description of the same intervention. Kazdin and Cole (1981) also investigated the effects of the label, content, and technical jargon on the teachers’ judgments related to behavior modification. The findings showed that humanistic and neutral procedures were perceived as more acceptable than behavior modification as such.

At the same time, Rhoades and Kratochwill (1992) examined teachers’ acceptance of behavioral consultation regarding consultant language and found no direct relationship between the consultant’s use of non-technical versus technical language on teachers’ treatment acceptability ratings. One possible explanation is that teachers are now becoming more familiar with behavioral techniques; thus, they have less bias against them.

**Contextual factors.** In addition to characteristics of the consumer and the intervention itself, *contextual factors* can influence treatment acceptability. The amount of time and resources available are contextual factors affecting treatment acceptability. Contextual factors may partly explain why teachers, despite positive perceptions of the applicability of evidence-based practices, seldom implement them in their classrooms (Burns & Ysseldyke, 2008). Studies have shown that the time and resources that an intervention requires can greatly influence its acceptability to teachers (Martens et al., 1985; Reimers et al., 1987).
Inadequate information about the practice and how to implement it are the other major impediments to the acceptability and use of evidence-based practices. When such information is limited or difficult to access, school staff will have problems implementing evidence-based practices. Many teachers cannot find the resources required to implement evidence-based practices because they do not have enough time to look for them or lack opportunities to access needed information (Powers, Bowen, & Bowen, 2010; Williams & Coles, 2007).

*Ecological intrusiveness* is another significant influence on teachers’ decisions to use or not use an intervention in the classroom. While ecological intrusiveness may be more likely to impact the treatment acceptability of behavioral interventions, it might also be a factor when academic interventions impact ongoing classroom routines. An ecologically intrusive intervention requires teachers to substantially alter their own behavior or the classroom ecology to bring about desired outcomes. For example, to increase desired behaviors, teachers may use more rewards than before, meaning that the ecological climate of the classroom is being affected by changes in teacher behavior. Such changes might also affect students who are not actually the target of the intervention (Witt, 1986), thereby leading to unexpected effects on other students’ behavior (Voeltz & Evans, 1982).

Additionally, dissemination of research findings has an important role in increasing the knowledge of teachers and preventing them from developing misconceptions with regard to specific interventions and their implementation in the classroom. Several studies have found that teachers who received in-service training in
ADHD were more likely to implement recommended practices than those who received little or no training (Martinussen, Tannock, & Chaban, 2011; Zentall & Javorsky, 2007). Also, Sciutto (2013) found a moderate relationship between strength of knowledge and strength of misconceptions. Tingstrom (1990) replicated Sciutto’s findings and stated that teachers who had knowledge of behavioral techniques had more positive perceptions of such techniques.

Issues in institutional support can become a barrier to the acceptability and implementation of reading interventions in classrooms, thereby increasing the gap between research and practice in special education. Powers and his colleagues (2010) indicated that one of the most significant barriers was failing to receive institutional support, training, and instruction about evidence-based practices. Sufficient teacher training and continuous assistance during the implementation process are clearly necessary in order to reach the expected achievement levels promised by evidence-based practices (Kutash et al., 2009).

My literature review indicates that there are many factors affecting treatment acceptability and the use or disuse of an intervention in classrooms. While all those factors have an important role for treatment acceptability, teacher factors could be considered as the critical one because teachers’ (consumers) judgments determine the level of treatment acceptability. Therefore, it is necessary to focus on which teacher factors affect teachers’ decisions of acceptability. Furthermore, previous studies mostly investigated treatment factors; however, little research has explored teacher factors. Also, the focus of those studies examining teacher factors were generally for behavioral
interventions. Thus, there is a need for examining teacher factors for academic interventions. Additionally, the development of measures of treatment acceptability provides one way to empirically investigate theoretical models of treatment acceptability and factors that influence treatment acceptability. In the next section, I provide an overview of treatment acceptability measures to identify how treatment acceptability is measured.

**Measurement Instruments to Assess Treatment Acceptability**

In this section, I first review four instruments that assess treatment acceptability for clinical use. Then, I review the instruments that measure the acceptability of school-based behavioral interventions and the acceptability of school-based assessment methods. I examined each instrument by considering the following elements: (1) why the instrument was developed, (2) how many items the instrument includes, (3) which treatment acceptability factors are identified, and (4) the reliability of the instrument.

My review of the literature indicates that treatment acceptability has generally been evaluated by administering a questionnaire. More specifically, respondents rated questions and/or statements using a Likert-type scale. The first instrument, the *Treatment Evaluation Inventory* (TEI) was developed by Kazdin in 1980 for clinical use. Witt and Martens (1983) also developed another measure, the *Intervention Rating Profile* (IRP) to assess treatment acceptability for school-based interventions. Both of these measurement instruments are commonly used in studies by making some modifications in the number of items and/or changing the wordings of the some items. This section provides a review of these measurement instruments.
Treatment Acceptability Measures for Clinical Treatments

Treatment acceptability measures that have been developed for (1) clinical treatments and (2) school-based interventions. My literature search revealed that there are four different measures that assess treatment acceptability for clinical use. These measures are the Treatment Evaluation Inventory, the Treatment Evaluation Inventory-Short Form, the Treatment Acceptability Rating Form, and the Treatment Acceptability Rating Form-Revised.

Treatment evaluation inventory. The TEI was designed for clinical use to measure the acceptability of treatments for children with behavior problems (Kazdin, 1980). Respondents complete 15 items using a 7-point Likert-scale that ranges from 1 (not at all acceptable) to 7 (very acceptable). Higher scores indicate that respondents’ judgments are more positive regarding the acceptability of the given interventions. The items assess participants’ acceptability of the treatment, willingness to implement the procedures of the treatment, appropriateness of the treatment procedures for the child’s other behavior problems, fairness of the treatment, and whether students like the procedures. To measure psychometric characteristics of the instrument, a pilot study was conducted on 60 undergraduate students from an introductory psychology class. Factor analysis indicated that the TEI had high loadings on one factor (acceptability) as 51.4% of the variance.

Treatment evaluation inventory-short form. While Kazdin’s (1980a) TEI was the first major measurement instrument of treatment acceptability, its length and terminology has limited its value in clinical research (Kelley, Heffer, Gresham, & Elliott,
1989). Therefore, Kelley and colleagues (1989) developed a short form of the TEI called the *Treatment Evaluation Inventory-Short Form* (TEI-SF). This instrument has 9 items with a 5-point Likert-type scale. Two experiments were administered. The first experiment was conducted to select items for the TEI-SF from the TEI. In this experiment, 153 parents completed the TEI. The factor analysis contradicted Kazdin’s (1980) study as two treatment acceptability factors were identified. Factor 1 was acceptability (42% of the variance) and Factor 2 was ethical issues/discomfort (19% of the variance). Items were selected by including above .78 correlations for the Factor 1 and .58 correlations for the Factor 2. The second experiment was conducted with undergraduate students (N= 264) from introductory psychology class. The purpose, again, was to investigate the factor structure and internal consistency of the TEI-SF. Results indicated that items loaded on two different factors; acceptability (57% of the variance) and discomfort (12% of the variance). Coefficient alpha (.85) that indicated the reliability of the instrument did not decrease by shortening and simplifying the items of the TEI.

**Treatment acceptability rating form.** The *Treatment Acceptability Rating Form* (TARF, Reimers & Wacker, 1988) is also a modified version of the TEI. The TARF was piloted and developed by administering it to 20 parents of children with behavior problems. The purpose of the TARF is to measure the acceptability of the effectiveness and cost of an intervention. The 15 items on the TARF use a 7-point Likert-type scale. Due to small sample size, the instrument was not subjected to factor analysis. The internal consistency range was 0.80 to 0.91.
**Treatment acceptability rating form-revised.** The *Treatment Acceptability Rating Form-Revised* (TARF-R, Reimers, Wacker, & Cooper, 1991) was a modification of the TEI designed to increase applicability to clinical populations. The instrument includes 20 items with 7-point Likert-type scale. Seventeen items of the scale were divided into six categories or variables; reasonableness, effectiveness, side effects, disruptiveness, cost, and willingness. The remaining three items of the scale address problem severity and understanding of the treatment. Ninety parents of children with behavior problems were included to the study. Cronbach’s alpha coefficient ranged from .90 to .92 for *total acceptability* score, .86 to .91 for *reasonableness*, .80 to .91 for *effectiveness*, .55 to .68 for *side effects*, .62 to .68 for *disruptive/time*, .42 to .68 for *cost*, .79 to .91 for *willingness*, and .91 to .96 for *problem severity*. The Cronbach’s alpha coefficient range was not reported for *understanding*.

My literature review indicated that the TEI is the first instrument for assessing treatment acceptability for clinical use. The other three measures are modified versions of the TEI. These modifications were made to decrease the number of items in the TEI to increase its value for clinical research, to measure additional factors (i.e., effectiveness and cost), or to change the wording to increase its clinical use. While general acceptability is the common factor for all the measures, there are also additional factors were found for the TEI-SF and TARF-R.

**Treatment Acceptability Measures for School-Based Interventions**

While the TEI and its modifications were developed to assess the acceptability of clinical treatments, it is necessary to address the acceptability of school-based
interventions. My search of the literature indicated that there are eight different measures that assess the acceptability of school-based interventions; the Intervention Rating Profile, the Intervention Rating Profile -15, the Abbreviated Acceptability Rating Profile, the Children’s Intervention Rating Profile, the Behavior Intervention Rating Scale, the Assessment Rating Profile-Revised, and the Curriculum Based Measurement-Acceptability Scale.

**Intervention rating profile.** The Intervention Rating Profile (IRP) was developed to assess the acceptability of classroom-based interventions. The IRP includes 20 items with Likert-type ratings that range from 1 (strongly disagree) to 6 (strongly agree). The IRP assesses (1) whether an intervention is acceptable for implementation in general education classrooms; (2) whether an intervention requires too much teacher time; (3) whether an intervention exposes the target child to risk or adverse side-effects; (4) whether the treatment negatively influences students in the classroom who are not targets of the intervention; and (5) whether the treatment is easy for a general education teacher. Preservice students and teacher candidates (N=180) participated in the study. Factor analysis determined that there was one primary factor along with four secondary factors that accounted for variance in treatment acceptability. The IRP had high loadings on the primary factor (*general acceptability*) which accounted for 41% of the variance. The loadings on the secondary factors were as follows: risk for the target child (9% of the variance); time for teachers to implement the treatment (9% of the variance); effects on other children (7% of the variance); and skills required for teachers to use the treatment (7% of the variance). Coefficient alpha of the instrument was reported as .91 (Witt &
Martens, 1983).

**Intervention rating profile-15.** The Intervention Rating Profile-15 (IRP-15, Martens et al., 1985) is a shortened form of the IRP. The IRP was revised to increase its internal consistency by adding eight items to the scale and deleting other items. The revised IRP includes 15 items using a 6-point Likert-type scale. General education and special education teachers (N= 54) were included in the development of the IRP-15. According to the results of a principal component analysis, the IRP-15 had one primary factor that is *general acceptability*. The item loadings ranged from .82 to 95. Cronbach’s alpha was reported as .98 (Martens et al., 1985).

**Abbreviated acceptability rating profile.** The Abbreviated Acceptability Rating Profile (AARP, Tarnowski & Simonian, 1992) is also a modified version of the IRP. The AARP was developed to address limitations of the IRP-15, such as item complexity and the administration time (Tarnowski & Simonian, 1992). The IRP-15 was modified in two ways: (1) some of the items were removed to improve content validity, and (2) the wording was changed to make items clearer and more readable. The AARP includes 8-items using a 6-point Likert-type scale. The AARP was initially administered to 60 mothers of children making pediatric outpatient visits. A cross-validation to replicate the results of the initial study was conducted by using a different sample of 80 parents (Tarnowski & Simonian, 1992). Last, the factors of *time-intensiveness* and *item readability* were evaluated by including a third independent sample of parents (N=30). According to the results of the initial sample, all items loaded on one unitary factor of *acceptability* (84.9 % of the variance). Internal consistency was examined by calculating
correlation coefficients between items of the scale and the total score of the AARP. The results indicated that there were positive significant correlations for all the items ranging from .88 to .96. Split half coefficient was .95 and Cronbach’s alpha was .97, indicating good internal consistency or reliability. Also, the cross-validation results indicated that all items loaded on a unitary factor (acceptability) which accounted for 90.3% of the variance in scores. Item correlations of the last sample ranged from .89 to .98. Split half coefficient was .97; Cronbach’s alpha was .98. Data from the last sample indicated that the AARP was completed in 10 minutes, which is the one-half time necessary to complete IRP-15. The Harris-Jacobson Wide Range Readability Formula indicated that the readability of the items for the AARP was 5.0.

**Children’s intervention rating profile.** The *Children’s Intervention Rating Profile* (CIRP, Witt & Elliott, 1985) was developed to assess children’s perceptions of behavioral interventions that are used by teachers. It is a modified version of the IRP. It has 7 statements using a 6-point Likert-type scale. The instrument was administered to 79 children. A factor analysis indicated that the CIRP had loadings on one primary factor (79% of the variance). One item (teacher harshness) did not have significant loadings on the primary factor. The coefficient alpha .89 (Elliott et al., 1986).

**Behavior intervention rating scale.** The *Behavior Intervention Rating Scale* (BIRS, Von Brock & Elliott, 1987) was developed to investigate the relationship between teacher effectiveness and the acceptability of interventions along with their relationship. The items on this scale were derived using a modification of the IRP-15. In addition to the 15 items on the IRP-15, the BIRS has 9 items that further assess treatment
effectiveness. These items were included under a subscale, the Effectiveness Rating Profile. Also, the Semantic Differential (Osgood, Suci, & Tannenbaum, 1957) was incorporated for the validity issues. The BIRS has 24 statements with 6-point Likert-type scales. The instrument was completed by 216 teachers attending graduate classes.

Findings indicated a high correlation ($r = .79$) between intervention effectiveness and intervention acceptability. Consequently, a nonorthogonal and a varimax rotations were conducted. The results showed that there were three factors: “acceptability, effectiveness, and time of effectiveness” (Von Brock & Elliott, 1987; p. 136). Factor loadings were as follows: acceptability, 63% of the variance; effectiveness, 6% of the variance; and time of effectiveness, 4.3% of the variance. The reliability of the each scale was found by calculating coefficient alphas indicating that the acceptability factor yielded alpha of .97, the effectiveness factor yielded alpha of .92, and the time to effectiveness factor yielded alpha of .87.

**Assessment rating profile.** The Assessment Rating Profile (ARP) was developed by Kratochwill and VanSomeren (1984). This instrument is also a modification of the IRP. The ARP has 18 items on 6-point Likert-type Scale. Some IRP items were excluded from the ARP. Also, some items were added and reworded. The psychometric characteristics of the ARP were first reported by Shapiro and Eckert (1994). The instrument was administered to 500 school psychologists. A factor analysis indicated that a factor loading on overall acceptability as 54.9% of variance. The loadings on the second factor (intrusiveness) accounted for 10.2% and the third factor (danger or high risk) accounted for 6% of variance. Cronbach’s coefficient alpha was calculated as .94.
**Assessment rating profile-revised.** *The Assessment Rating Profile-Revised* (ARP-R) was developed by Eckert, Hintze, and Shapiro (1999) to address weaknesses of the ARP. For the ARP, a principal components analysis conducted which caused to have the fragmented factors and weak loadings on the factors. Also, because of the many weak factor loadings, removing poor factor loaded items and changing wording of some items were thought to provide a better instrument in terms of reliability and validity (Eckert et al., 1999). Thus, for the ARP-R, three items were deleted because of poor factor loadings. Also, two more items were removed because of high loadings on the secondary factor. Last, an irrelevant item was deleted from the instrument. The ARP-R has 12 statements on 6-point Likert-type scale. The instrument was completed by 201 general education teachers. Confirmatory factor analyses indicated that the ARP-R has sufficient psychometric properties. The Cronbach’s coefficient alpha was .99.

**Curriculum based measurement-acceptability scale.** Another instrument is the *Curriculum Based Measurement-Acceptability Scale* (CBM-AS) as developed by Oats and Allinder (1995). This scale measures teachers’ judgments of curriculum-based measurement (CBM). The instrument has 20 statements and uses a 6-point Likert-type scale, ranging from 1 (strongly disagree) to 6 (strongly agree). The instrument assesses the perceived effectiveness of the CBM, the acceptability of the time required to implement the CBM, and skills or training needed to use the CBM (Allinder & Oats, 1997). CBM-AS reliability indicated that Cronbach’s alpha was .90.

My literature search indicated that the IRP is the most frequently and widely used instrument to assess the acceptability of school-based interventions. Other instruments
are modifications of the IRP or IRP-15. These modifications were made to increase the internal consistency of the IRP, to decrease item complexity and the length of administration time of the IRP-15, to assess children’s acceptability of behavioral treatments, to assess additional factors (i.e., effectiveness and time of effectiveness), and to measure the acceptability of assessment methods. While these instruments vary regarding their purpose (e.g., measuring teachers’ treatment acceptability or children’s treatment acceptability), they were developed to measure the acceptability of behavioral interventions and the acceptability of assessment methods. My review of literature suggests that there is no measure that is specifically designed to assess the acceptability of reading interventions. On the other hand, some studies have investigated the acceptability of reading interventions. Therefore, it is beneficial to review these studies to identify which instruments were used to assess the acceptability of reading interventions.

Investigating the acceptability of reading interventions is important to the field because many factors that contribute to the research to practice gap and use of effective practices are either directly or indirectly related to treatment acceptability. No matter how effective the intervention is, implementing high-quality interventions in classrooms depends on teachers’ acceptance of the intervention procedures. Since there is a correlation between intervention acceptability and its use, the intervention is more likely to be implemented with fidelity when a treatment is perceived as acceptable (Reimers et al., 1992). In the next section, I review research on the acceptability of reading interventions and measures that are used to assess the acceptability of reading interventions.
Acceptability of Reading Interventions

In this section, I review studies (methods, measures, findings) that examined teachers’ acceptability of reading interventions to obtain teachers’ judgments toward reading interventions and instruments used to measure acceptability of reading interventions. My criteria to select the studies are as follows: (1) treatment acceptability must be a major dependent variable of the investigation, (2) acceptability of interventions for students with reading problems must be investigated, and (3) treatment acceptability needs to be assessed by an instrument specifically developed for assessing treatment acceptability.

There are many studies that investigate teachers’ attitudes, beliefs, and perspectives toward recommended practices (e.g., Deford, 1985; Boardman et al., 2005; Burns & Ysseldyke, 2008); however, the methodology of these studies is mainly based on interviews and/or researcher-created surveys. Additionally, there are some studies whose main purpose is to measure effectiveness of an intervention but at the same time gather participants’ treatment acceptability to assess social validity (e.g., Saenz, Fuchs, & Fuchs, 2005; Macaruso, Hook, & McCabe, 2006; Skindrud & Gersten, 2006). Since their main focus is not to measure treatment acceptability, limited information was included about the acceptability of given interventions; thus, they were not included in this review. Four studies met my criteria for inclusion in this review.

The studies related to acceptability of interventions for students who struggle in reading generally focus on acceptability of group contingencies and effects of group contingencies on students’ spelling performance. For example, the study by Turco and
Elliott (1990) was investigated the effectiveness of group contingencies on students’ spelling performance, students’ social status, and intervention acceptability. Seventy-four fifth grade students were selected as participants. A pretest-posttest control group design was employed. Students received spelling instruction in pre-treatment and post-treatment phases. Their achievement levels were tested by administering the Wide Range Achievement Test-Revised (WRAT-R, Jastak & Wilkinson, 1984) and the Basic Goals in Spelling (Kottmeyer & Claus, 1976). Also, the CIRP (Witt & Elliott, 1985) was conducted to see the students’ acceptability ratings of the intervention. The results of the study showed that there were no significant differences in students’ spelling performances meaning that various types of group contingencies did not affect the students’ spelling outcomes. Also, various group contingencies did not have significant effects on peer-nominated social status. Regarding the intervention acceptability ratings, a series of one-way ANOVA indicated that post-treatment phase was rated significantly less acceptable than pre-treatment phase in all treatment groups except the dependent individual group. Also, researchers found that interdependent group tasks were found significantly more acceptable than “the dependent groups, dependent individual, and no incentive individual methods” (Turco & Elliott, 1990, p.32).

Another study also examined children’s acceptability of intervention in group contingencies planned to increase spelling performance and investigated the influences of group largeness and independent contingencies on students’ spelling performance (Shapiro & Goldberg, 1990). Four intervention group combinations were conducted namely “independent-small, independent- large, dependent-small, and depended-large”
(Shapiro & Goldberg, 1990, p.236). Spelling instruction was provided to all students in the groups. Also, token economy system, which was not used to teach spelling and academic practices by teachers before, was conducted. The study included two different investigations. The purpose of the first investigation was to examine “the relationship of group size to interdependent and dependent group contingencies” (Shapiro & Goldberg, 1990, p.233). The first study involved forty-five elementary school children. An alternating crossover treatment design and a randomized assignment to the two intact groups were established. Subgroups within each group were also created as small size and large size. Students under each group received the spelling intervention. The 10-item spelling test and the CIRP (Witt & Elliott, 1985) were administered before and after the spelling intervention. The results of the study indicated that students in the low group whose performance below 80% improved their spelling performance from baseline to the first intervention phase. Also, the improvement of the students on spelling performance maintained through the second intervention phase. Spelling improvement was gained under the both groups: interdependent and dependent. Results indicated that there was no relationship between spelling accomplishment and group size. Also, no relationship was found between group size and kind of group contingency. After students receive the intervention, they were rated the acceptability of the group contingencies. The results of repeated measures ANOVA indicated that the post-treatment of all group contingencies was rated as more acceptable than the pre-treatment. The purpose of the second investigation was to examine “the effects of group size and interdependent contingencies only” (Shapiro & Goldberg, 1990, p.233). The second study involved forty-eight
elementary school children. An alternating crossover treatment design was conducted. Spelling test results and acceptability ratings were gathered as in the first study. The results indicated that students who have poor performance were significantly improved their spelling under the interdependent group contingency. In contrast to the first study, no differences were found between two group sizes: interdependent and dependent contingencies. Small group interdependent contingency was rated more acceptable than large group interdependent contingency. Similar to the first investigation, the acceptability ratings were higher for post intervention than the initial intervention.

Overall, the differences in-group sizes were not consistent across studies; however, researchers indicated that different group sizes could be helpful to increase students’ spelling performance. Also, finding no relationship between intervention acceptability ratings and intervention outcomes showed that other variables that might affect the relationship need to be examined to better explain the findings (Shapiro & Goldberg, 1990).

Mautone et al. (2009) investigated the relationship between intervention integrity and intervention acceptability for “reading interventions across two consultation models, intensive data-based academic intervention (IDAI) and traditional data-based academic intervention TDAI” (p. 919). Eighty-three students who have ADHD and problems in reading and/or math were selected as participants. Also, their special education \( (n=6) \) and/or general education teachers \( (n=66) \) were selected as participants. Matched experimental control group design was conducted. After completing the intervention condition, rating scales to measure intervention acceptability were provided to the
groups. The intervention acceptability scale was the BIRS (Elliott & Treuting, 1991). In order to measure acceptability of academic interventions, the BIRS was modified. Additionally, treatment integrity checklists were provided to the participants. The results indicated that there was a moderate positive relationship between intervention integrity and intervention acceptability for two consultation groups. For IDAI, treatment integrity was found significantly greater than TDAI as \( t[81] = 3.55; p < .01 \); however, the intervention acceptability did not significantly differ between the groups (\( t[81] = 0.75; p = .456 \)).

Last, the study by McCurdy and Shapiro (1992) investigated the effects of four types of progress monitoring on the rate of children’s oral reading. These progress-monitoring forms were teacher-monitoring, peer-monitoring, self-monitoring, and no monitoring. Then, students’ and teachers’ acceptability of these four monitoring forms were gathered. Forty-three elementary school children who have learning disabilities in reading were selected as participants. Also, ten teachers were selected from the resource room to monitor students’ reading rate progress. Students were randomly assigned to the four types of progress monitoring conditions. Number of students assigned to the conditions as follows: teacher-monitoring, \( n=10 \); peer-monitoring, \( n=10 \); self-monitoring, \( n=12 \); and no monitoring, \( n=11 \). Student progress was measured two times a week for nine weeks. The results of the idiographic comparison indicated that students in the self-monitoring condition had the largest improvement on their reading rate while student gains in the teacher-monitoring condition were consistent across time. Teachers’ acceptability rates were measured by a modified version of the IRP-15 (Martens et al.,
1985). Also, students’ acceptability rates were measured by a modified version of the CIRP (Witt & Elliott, 1985). The results of a series of repeated-measures multivariate analysis of variance (MANOVA) showed that there were differences among the ratings of three conditions: teacher-monitoring, peer-monitoring, and self-monitoring. Ratings of students in the self-monitoring condition indicated that self- and peer-monitoring conditions were significantly more acceptable than teacher-monitoring condition as \( p < .02 \). On the other hand, repeated measures ANOVA indicated that there were no significant effects for treatment or time regarding teachers’ responses on the IRP-15. The study suggested that peer- and self-monitoring procedures were beneficial for special education teachers in decreasing the amount of time spend to administer progress monitoring in the classroom.

Reviewed literature indicated that studies examined the acceptability of reading interventions focused on the acceptability of some specific academic interventions such as spelling instruction in different group contingencies, reading interventions across two consultation models, and four types of progress monitoring. While three of them examined children’s acceptability of reading interventions (e.g., Shapiro & Goldberg, 1990; Turco & Elliott, 1990; McCurdy & Shapiro, 1992), two of them examined teachers’ acceptability of academic interventions (e.g., Mautone et al., 2009; McCurdy & Shapiro, 1992). To measure the acceptability of these interventions, researchers administered the modified versions of the IRP-15, the CIRP, and the BIRS. The researchers basically changed the wordings of the instruments and used them in their studies; however, none of the studies examined the psychometric characteristics of the
modified versions of the instruments. As stated previously, literature shows that there is no measure exist to assess the acceptability of reading interventions. On the other hand, to have reliable and consistent research results, it is necessary to use an instrument that is specifically developed for measuring the acceptability of reading interventions and that has psychometric properties.

**Purpose of the Study**

To date, research has focused primarily on the treatment acceptability of behavioral interventions while very little research has investigated the treatment acceptability of academic interventions. Also, literature review indicated that many factors affect teachers’ treatment acceptability. Even though each of those factors is a significant part of the issue of treatment acceptability, teacher factors need to be considered as the critical one because teachers’ judgments, as consumers, determine the acceptability of effective interventions. Specifically, very little research has explored teacher factors influencing treatment acceptability. Also, those studies investigating teacher factors generally focused on behavioral interventions. Thus, there is a need for examining teacher factors for reading interventions. Furthermore, no instruments have been developed to assess the treatment acceptability of reading interventions. The development of measures of treatment acceptability provides one way to empirically investigate the treatment acceptability of reading interventions and examine factors that influence treatment acceptability. In this study, I empirically test and develop a conceptual model of treatment acceptability of reading interventions by developing a reading intervention treatment acceptability rating scale.
The first purpose of this study is to develop and investigate, using exploratory factor analysis and other statistical analyses, a new instrument I created called the Reading Intervention Rating Profile (RIRP). The RIRP is designed to measure teachers’ acceptance of reading intervention. Because the focus is on treatment acceptability, the RIRP is designed to assess the acceptability of intervention procedures (as opposed to intended and/or actual intervention outcomes), as one critical part of an overall model of social validity. The second major purpose of my study is to investigate whether teachers’ acceptability of reading interventions differ by teacher demographics (i.e., teacher certification type, level of experience, level of education).
CHAPTER 2

METHOD

Overview

Empirically validated reading interventions are necessary for students who exhibit significant reading deficits. As indicated in the chapter one, the purpose of this study was to develop and conduct an initial investigation of a measure designed (Reading Intervention Rating Profile, RIRP) to assess teachers’ acceptability of reading interventions and to investigate whether teachers’ ratings on the RIRP differ depending on their characteristics. In this chapter, I outlined the methods used to conduct this study, including 1) the research questions; 2) sampling procedures for determining the necessary sample size for the planned analyses; 3) the procedures for developing the new measurement instrument; 4) a description of the analyses; and 5) a description of the research design. I also provided a plan for how data was analyzed.

Research Questions

The purpose of this study was to answer the following research questions:

1. What are the psychometric properties of the RIRP?
   a. What is the factor structure of the RIRP for special education teachers?
   b. What is the factor structure of the RIRP for general education teachers?
   c. What is the reliability of the RIRP for special education teachers?
   d. What is the reliability of the RIRP for general education teachers?

2. Do the acceptability ratings of teachers on the RIRP vary depending on certain teacher demographics?
a. Do the acceptability ratings of special education teachers on the RIRP differ by their teaching experience (i.e., high experienced or low experienced)?
b. Do the acceptability ratings of general education teachers on the RIRP differ by their teaching experience (i.e., high experienced or low experienced)?
c. Do the acceptability ratings of special education teachers on the RIRP differ by their level of education (i.e., bachelor’s degree or masters degree)?
d. Do the acceptability ratings of general education teachers on the RIRP differ by their level of education (i.e., bachelor’s degree or masters degree)?

Since the factor structure of the RIRP for special education teachers was found to be similar to the factor structure of the RIRP for general education teachers, the following research question was also examined.

e. Are the acceptability ratings of special education teachers on the RIRP different than general education teachers?

Sampling Procedures

The sampling procedures including determination of sample size and recruitment are explained as follows.

Determination of Sample Size

For the initial examination of a measurement instrument, the recommended approach is to have at least 1:10 item-to-respondent ratio to obtain a reliable factor structure (Harman, 1976; Schwab, 1978). Considering these recommendations, if an instrument has 40 items, at least 400 respondents are necessary to analyze reliability and to examine the factor structure of the instrument. Also, when conducting an exploratory
factor analysis (EFA), guidelines suggest that a sample of 150 respondents is sufficient (Hinkin, 1998) without considering the number of items. Therefore, based on these recommendations and considering the total number of items in the newly developed scale ($N = 15$), the goal was to recruit at least 150 special education teachers and 150 general education teachers to participate in the study, for a total of 300 participants.

**Recruitment**

Participants were general education teachers and special education in elementary grades from first grade through fifth grade. In this study, only elementary school teachers were selected as participants. One question in the demographics section of the survey asks participants how many years of teaching experience they have with struggling readers. Based on their answers, only teachers who have experience with struggling readers were selected. Struggling reader can be defined as any student below average in reading. There was a reason of selecting only teachers who have experience with students with reading problems: Participants’ grade levels and experience with struggling readers needed to match with the student’s characteristics in the case scenario to figure out whether the given intervention was appropriate and acceptable for that specific student. Therefore, to prevent any confusion, teachers who did not have any experience with struggling readers, middle school teachers, and high school teachers were not be selected as participants for the study.

For the purpose of the analyses planned for this study, there was no upper limit regarding the number of respondents. Therefore, “snowball” sampling was considered (Goodman, 1961) prior to using a purposive sampling. “Snowball” sampling allows
researchers to reach potential respondents by asking them to forward and share the survey with other elementary school teachers. This “snowballing” approach allows researchers to recruit participants who satisfy study requirements (Biernacki & Waldorf, 1981). This process continues until a sufficient number of participants for the planned analyses is obtained. However, I was not able to ask teachers to share the survey with other elementary school teachers because I would not be sure if I had permission from the district administrators and/or principals of those who the survey was going to be shared.

Convenience sampling is a kind of nonrandom sampling, was also considered because it allows researchers to select participants according to their proximity (accessibility), availability, and willingness to be part of the study (Kitchenham & Pfleeger, 2002). The main purpose of convenience sampling is to gather information from the members of the target population based on practical criteria (Etikan, Musa, & Alkassim, 2016). However, due to several limitations of the convenience sampling (i.e., limited generalizability, having poor quality data) (Etikan et al., 2016), purposive sampling was administered, which allows researchers to collect participants according to predefined criteria.

Purposive sampling was selected to conduct because the participants of the study have some specific characteristics (i.e., having experience with struggling readers, currently working at elementary schools). Purposive sampling is also a nonrandom technique that helps researchers get relevant information for the study by selecting participants based on specific criteria. Also, availability and willingness of the participants are important elements of purposive sampling (Etikan et al., 2016). While the
School districts in a Midwest state were contacted about the current study. Emails were sent to 574 school districts administrators to have permission to conduct the study in their district. A brief written overview of the study purpose and procedures were provided to the administrators. 80 school districts approved to participate in the study. With the permission from district administrators, e-mails were sent to 173 elementary school principals requesting their permission to contact with their teachers in their school. Some school principals provided the email addresses of their teachers to allow me to directly contact with the teachers. After having permission from the school principals, reaching out teachers to participate in the study occurred in two ways: (1) an email with the survey link and a consent form (see Appendix A) was distributed by school district administrators and/or school principals; and (2) an email with the survey link and the consent form was sent directly to teachers’ email addresses by me as a researcher to request their participation in the study. After one month, reminder e-mail was sent to the participating districts and schools to complete the survey.

**Development of the RIRP**

A systematic review of the literature, described in the chapter one, revealed that researchers have not yet examined the psychometric properties of an instrument that measures teachers’ acceptability of reading interventions. In this section, I describe in
detail the steps I completed to modify an existing measure of treatment acceptability (the Intervention Rating Profile-15) to create a measure, the *RIRP*, specifically designed to measure the treatment acceptability of reading interventions. This study provided the initial evaluation of the *RIRP* by examining the psychometric properties of the measurement (see Appendix B).

**The Reading Intervention Rating Profile**

The *RIRP* is a modification of the Intervention Rating Profile (Martens et al., 1985), which assesses teachers’ judgments regarding the acceptability of behavioral interventions. The *RIRP* was designed for assessing teachers’ acceptability of reading interventions. The measure can be completed either in online or hardcopy format. It includes 15 items asking teachers to rate reading interventions using a 6-point Likert scale (i.e., 1 = Strongly Disagree, 2 = Disagree, 3 = Slightly Disagree, 4 = Slightly Agree, 5 = Agree, 6 = Strongly Agree). Strongly Agree indicates that respondents have a positive perception of the treatment. The items inquire about teachers’ judgments of the intervention procedures (i.e., appropriateness, fairness, effectiveness, and benefits to the students with reading problems). The range of the total score is 15 to 90 on the 15-item *RIRP*. A higher score indicates higher treatment acceptability. To modify the IRP-15, I followed systematic modification procedures. These procedures are explained below.

**Item development of the RIRP.** Several steps followed to develop the *RIRP* to maximize its reliability and validity. The first step was to review the literature to investigate if there exist measures of teachers’ acceptability of reading interventions. After the literature review, several instruments were found (e.g., the TEI, the IRP-15, the
BIRS, the CIRP). They were developed to measure treatment acceptability; however, none of these instruments was developed to measure the treatment acceptability of reading interventions.

Since there was no instrument that measures teacher acceptance of reading interventions, the second step was to select an instrument to modify by considering three criteria. The instrument and scores produced from the instrument need to be (a) developed for school-based interventions, (b) applicable to teachers, and (c) psychometrically sound (i.e., adequate reliability and validity). According to these criteria, the IRP-15 was a good candidate for modification. First, the IRP-15 includes items to assess the acceptability of school-based interventions. Second, the IRP-15 is the only instrument that examines teachers’ perspectives regarding interventions. While there are other measures that also assess teachers’ treatment acceptability, they are a modification of either the IRP (e.g., AARP) or the IRP-15 (e.g., BIRS).

Regarding the psychometric properties, the IRP-15 has one general acceptability factor and item loadings ranged from .82 to .95. On the other hand, the IRP has one general acceptability factor (41% of the variance) and several secondary factors discussed previously in the literature review section. Also, while Coefficient Alpha (i.e., internal consistency reliability) for the IRP-15 was reported as .98 (Martens et al., 1985), it was reported as .91 (Witt & Martens, 1983) for the IRP. These results indicate that both the IRP-15 and the IRP have high internal consistency; however, the IRP has 20 items and the IRP-15 has 15 items. Choosing the scale with fewer items is more efficient because its internal consistency remains high with a smaller scale. Therefore, I have selected the
IRP-15 to modify for the purpose of developing a measure of teachers’ acceptability of reading interventions. Because the IRP-15 is based on the acceptability of behavior interventions, the wording of items was changed to focus on reading interventions.

**Demographic questions.** Along with the 15 *RIRP* items, the survey included 10 demographic questions. The demographic survey questions included teachers’ age, gender, race/ethnicity, level of education, years of teaching experience, experience with struggling readers, type of teaching certification, grade/grades currently teach, and reading programs and/or instructional approaches currently used with their students. It also asked teachers whether they use response to intervention (RtI) in their schools (see Appendix C). While some of the demographic questions (e.g., years of teaching experience, certification type, level of education) were directly related to the research questions, some questions (e.g., experience with struggling readers, grade/grades currently teach) were included to the demographics due to participant selection purposes, as only teachers who are currently working and have had experience with struggling readers were selected as participants. Additionally, remaining questions were included to the demographic portion to understand issues behind teachers’ responses (e.g., age, gender, race/ethnicity, reading programs and/or instructional approaches currently used, use of RtI).

The demographic portion of the rating scale included a total of ten questions requiring 2 to 3 minutes to complete. The case scenario and the 15 items of the *RIRP* were completed before the demographic section.
**Case scenario.** Prior to completing the *RIRP* rating scale, participants were asked to read a case scenario describing the characteristics of a fictional student (“Danny”) exhibiting reading problems. Specifically, Danny experiences difficulties in reading fluency (the ability to read text accurately, quickly, and with expression). Participants were told that Danny’s reading fluency is significantly below grade level. Additionally, a description of a reading intervention (repeated reading) selected by Danny’s teacher was provided, as well as steps for teacher implementation (see Appendix D). Repeated reading is a fluency intervention supported by research (Alber-Morgan, 2006; Samuels, 1979; Therrien, 2004). Repeated reading was selected because it can be implemented with students from elementary school to high school (e.g. Begeny, Krouse, Ross, & Mitchell (2009); Devault & Joseph, 2004; Fuchs, Fuchs, Hosp, & Jenkins, 2001; Huang, Nelson, & Nelson, 2008; Lo, Cooke, & Starling, 2011; Musti-Rao, Hawkins, & Barkley, 2009; Nelson, Alber, & Gordy, 2004; Oddo, Barnett, Hawkins, & Musti-Rao, 2010; Paige, 2006; Therrien & Hughes, 2008; Vandenbergen, Boon, Fore III, & Bender, 2008). Furthermore, repeated reading interventions are easy to understand and have simple steps to follow (Huang et al., 2008).

The aim of the repeated reading strategy is to increase the number of words that the student reads accurately and with expression in a specific time period (e.g., one minute; Samuels, 1979). While the student reads the text aloud, the teacher starts the timer. If the student hesitates to read a word for longer than five seconds or misreads the word, the teacher reads the word aloud and prompts the student to repeat the word correctly. If the student asks for help reading a word, the teacher reads the word aloud or
provides the word’s definition (WWC, 2014). This process continues two or more times until the student reaches a fluency criterion (words read correctly per minute; Therrien, 2004). Repeated reading sessions are typically administered daily at least three times a week (Alber-Morgan, 2006). Each session takes fifteen to twenty-five minutes and sessions continue throughout the year.

In the case scenario, the procedures of the repeated reading intervention were described step-by-step. The intervention had seven steps and included direct explanation of the lesson rules, modeling, leading, and practice. The intervention also consisted of error correction and praising procedures. The specific, conventional name used for the intervention (repeated reading) was omitted from the case scenario. I did not include the name because I did not want participants’ ratings to be influenced by familiarity with the name of the intervention.

Validity

To establish the content validity of the RIRP rating scale, I used expert reviews. Revisions were made based on the experts’ feedback and suggestions. Expert reviews included following steps.

**Expert review.** Content validation measures whether the content (items) of an instrument sufficiently illustrate the construct of particular interest (Crocker & Algina, 1986). It shows how certain people perceive the content of a measurement instrument in terms of reflecting a proposed goal efficiently (Shavelson, 1988). A typical process for establishing content validity is to have independent judges who find out whether the items in the instrument adequately reflect the construct of specific interest (Anderson &
Prussia, 1997). The judges need to be experts on subject matter and they are informed to examine the instrument systematically (Crocker & Algina, 1986). Therefore, three expert opinions and views were gathered to examine and enhance content validity. A written copy of the RIRP and a literature review of the issue of treatment acceptability and an evaluation sheet (see Appendix E) were distributed to the experts. They were asked to review the measure to indicate (1) how the items are relevant to the content (2) how clear the instructions, wordings, and the use of English, and (3) how the items need to be rephrased if necessary (Fowler, 2002). The experts were also asked to critique the RIRP in general and to provide their suggestions regarding excluding the items or including new items to the instrument. Prior to conducting the experiment, the judges’ suggestions were considered and some items were rewritten.

**Data Collection**

Data collection occurred online. There are many reasons to use an online survey. First, the number of participants needed to conduct the analyses was large and requires reaching teachers in many elementary schools. One of the advantages of administering surveys online is the ability to distribute the surveys quickly to many people. Participants are able to complete the online survey and return it at their convenience (Sax, Gilmartin, & Bryant, 2003; Sue & Ritter, 2012). Administering surveys through a paper-and-pencil format, on the other hand, limits the reach of a survey study, in terms of both time and potential respondents (Greenlaw & Brown-Welty, 2009).

Also, conducting surveys online reduces costs associated with printing and mailing (Clayton & Werking, 1998; Cobanoglu, Warde, & Moreo, 2001) and increasing
the sample size does not affect the cost (Benfield & Szlemko, 2006). Once participants are able to access a network, collecting data electronically reduces research cost as compared to sending questionnaires through the mail (Kiesler & Sproull, 1986). Therefore, online administration was an appropriate method since resources for conducting research (e.g. funding) were limited.

Regarding the quality of the data, some researchers discussed that the items of the online surveys are more likely to be skipped compared to traditional surveys (Wyrick & Bond, 2011) and there might be differences between online surveys and traditional surveys in terms of participants’ responses (Wright, 2005). However, it is difficult to assess whether the differences are due to the participants’ deceit or due to the differences between online and paper surveys (Wright, 2005).

Last, data collection in online format saves time when entering data (Ary, Jacobs, Sorensen, & Razavieh, 2010) and help minimize entry mistakes such as missing and inconsistent data (Touvier et al., 2010). While data are entered manually when traditional paper surveys are used, digital data entering occurs as the respondents complete the survey online (Touvier et al., 2010), decreasing data entry time and mistakes.

Survey Monkey® is an online survey tool that can be used to administer online surveys to large numbers of participants. For my study, Survey Monkey® was used to upload, distribute, and retrieve surveys. A Survey Monkey® link to the survey (including consent form) was emailed to the participants (see description of recruitment procedures, above). The consent form included the purpose and procedures of the study. Also, the information about the confidentiality of the survey and voluntary participation were
provided to the participants. Each participant was asked to answer the questions of the demographic survey, to read the case scenarios, and to evaluate the given intervention by completing the *RIRP*. To increase response rates, participants had a chance to win a $50 gift card to Amazon.com. Two participants had an opportunity to receive $50 Amazon gift card. By completing this survey, participants who provide their e-mail addresses were eligible to enter into prize drawing for the $50 Amazon gift cards.

**Research Design**

In this study, I first conducted an initial evaluation (exploratory factor analysis) of the Reading Intervention Rating Profile, a modified rating scale for measuring teachers’ treatment acceptability of reading interventions. Next, I examined whether the teachers’ ratings of the treatment acceptability of reading interventions using the *RIRP* vary by teacher characteristics. The Statistical Package for the Social Science (SPSS) was used to analyze the data. Descriptive analyses and three different statistical tests were conducted to answer the research questions.

First, for research questions 1A and 1B, two separate exploratory factor analyses (EFA; described below) were conducted to investigate the factor structure of the *RIRP* for special education teachers and general education teachers. Regarding EFA procedures, first, the assumptions of factor analysis were tested. Then, a method of factor extraction was selected. Last, factor rotation procedures were discussed. Also, Cronbach’s (Coefficient) Alpha was used to assess the internal consistency reliability of the scores on the measure. For the remaining research questions, one independent sample t-test and four Mann–Whitney *U* Tests were conducted. For research question 2A, an independent
samples t-test was administered to examine whether there is a difference between the ratings of low experienced special education teachers and high experienced special education teachers. For research question 2B, research question 2C, research question 2D, and research question 2E, the most appropriate statistical technique to use was the Mann-Whitney $U$ Test. Specifically, the Mann-Whitney $U$ Test was administered for research question 2B to examine whether there is a difference between the ratings of low experienced general education teachers and high experienced general education teachers. For research question 2C, the Mann-Whitney $U$ Test was used to assess whether there is a difference in the scores of special education teachers who have different level of education. Also, for research question 2D, the Mann-Whitney $U$ Test was used to assess whether there is a difference in the scores of general teachers who have different level of education.

The factor structure of the RIRP for special education teachers was found similar to the factor structure of the RIRP for general education teachers. Thus, research question 2E was also considered. In this case, the Mann-Whitney $U$ Test was administered to find out if there is a difference between special education teachers’ and general education teachers’ ratings on the RIRP. These analyses are explained as follows.

**Exploratory Factor Analysis Procedures**

In quantitative research, factor analysis is a multivariate statistical procedure (Thompson, 2004) that is commonly used in psychological and educational research (Hogarty, Hines, Kromrey, Ferron, & Mumford, 2005). Factor analysis is conducted for three reasons: (1) to figure out the structure of latent variables (also known as factors);
(2) to develop a measure of a particular construct; and (3) to combine and reduce large
data sets that have many variables to have more manageable data set (Yong & Pearce, 2013). There are two types of factor analysis: (1) exploratory factor analysis (EFA), and
(2) confirmatory factor analysis (CFA). EFA is generally preferred when researchers do
not know how many factors are required to clarify the interrelationships among a set of
items (Pett, Lackey, & Sullivan, 2003). CFA, however, is used to assess whether or how
well a hypothesized model of factors and their interrelationships fits the data (Pedhazur &
Schmelkin, 1991). In other words, while EFA helps researchers to generate the principal
features of a model from a large set of variables (Henson & Roberts, 2006), CFA helps
researchers test a proposed model or a priori theory by examining how well data fit the
model or theory (Williams, Onsman, & Brown, 2010).

The first research question of this study has four components: Research question
1A, research question 1B, research question 1C, and research question 1D. Research
question 1A and research question 1B focus on the factor structure of the RIRP.
Researchers use EFA to determine how many factors are required to clarify the
interrelationships among the items (Pett et al., 2003). Therefore, to answer research
question 1A and research question 1B, the EFA was used to explore the factor structure
of the RIRP and to generate a model for teachers’ acceptability of reading interventions.

Prior to investigating the factor structure, the following assumptions of factor
analysis were tested to ensure that the data are suitable to administer an EFA. First, the
sample size must be adequate to conduct factor analysis. As discussed previously, Hinkin
(1998) suggested that a sample of 150 respondents to a measure is adequate for EFA.
Also, another recommendation is called “sample to variable ratio” that is necessary to determine sufficient sample size (Hogarty et al., 2005). It is suggested that at least 1:10 (item: sample) ratio is necessary to be able to establish a reliable factor structure (Yong & Pearce, 2013).

The second assumption I assessed was the linearity of the variables. Because factor analysis is based on the assumption of linear relationships, the values must be linear (Gorsuch, 1983). To check whether there are linear relations among variables, scatterplots were examined. Also, data need to be appropriate for correlation analysis indicating that variables need to be at least interval (Floyd & Widaman, 1995). The RIRP includes 6-point Likert- scale items and is considered as an ordinal scale; however, Lubke and Muthén (2004) demonstrated that it is possible to evidence true parameter values in factor analysis with Likert scale data if the assumptions are met. On the other hand, applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons (Lubke & Muthén, 2004).

Third, the assumption of multicollinearity was tested. Multicollinearity means high correlations among independent variables which cause high sampling errors of partial correlations (Blalock, 1963). In factor analysis, the statistical tests must rule out significant multicollinearity, meaning that each variable in the data set must be unique (Tabachnick & Fidell, 2007). To assess multicollinearity, the correlation matrix (Factorability of R) was examined. Factorability for correlation coefficients needs to be over 0.3. Factorability of $R \pm 0.3$ indicates minimal loadings, $\pm 0.4$ indicates important loadings, and $\pm 0.5$ indicated practically significant loadings (Hair, Anderson, Tatham, &
Black, 1984). Additionally, Kaiser-Meyer-Olkin measure of sampling adequacy (KMO; Kaiser, 1974) and Bartlett’s test of sphericity (Bartlett, 1950) will be administered to determine the factorability. For factor analysis, the results of KMO need to be equal to or higher than .50 (index ranges from 0 to 1) to be acceptable and the results of Bartlett’s Test of Sphericity should be significant as $p < .05$ (Williams et al., 2010).

After testing the assumptions, EFA was conducted. The analysis included factor extraction and factor rotation procedures. I explain these procedures below.

**Factor extraction.** To conduct EFA, a method of factor extraction must be selected. Then, procedures to determine the number of factors are examined. There are many ways to extract factors such as principal component analysis (PCA), principal axis factoring (PAF), maximum likelihood, and image factoring. In the literature, PCA and PAF are commonly used by researchers (Williams et al., 2010). The main difference between these two extraction methods is the way that the communalities are used. The communality of an item means the explained variance by the factors that are going to be extracted. The estimation of communalities can range from 0 to 1.00. Higher values show higher explained variance of an item by the extracted factors (Pett et al., 2003).

PCA is preferred by some researchers because it is straightforward and easy; however, “it does not separate out errors of measurement from shared variance” (Pett et al., 2003, p. 102). While PCA assumes that the error variance is 1, meaning that there is no error variance, PAF assumes error variance (Rietveld & Van Hout, 1993). This renders PAF more conservative than PCA (Williams et al., 2010). Also, PCA is typically used as a data reduction technique (Osborne & Costello, 2009).
PAF was used in my study because it estimates correlations, and it does not include measurement errors (Pett et al., 2003). Also, PAF is recommended specifically when there are fewer than 35 variables (Velicer & Jackson, 1990). In this study, PAF was appropriate for three reasons: (a) the RIRP is not a perfect measurement in terms of reliability and it is assumed that there is error variance; (b) the number of variables of the RIRP are fewer than 35 ($N = 15$); and (c) this method produces more firm and replicable results and finds out “any underlying structure caused by latent variables” (Osborne & Costello, 2009, p. 133).

To select the number of factors to extract, it is suggested to use multiple criteria (Hair et al., 1984). One way to select the number of factors is to use a scree plot test. Therefore, a scree plot was examined via the graphical presentation to determine if the eigenvalues are large enough (Cattell, 1966). Since examining a scree plot is believed by some researchers to be too subjective (e.g., Gorsuch, 1983; Thompson, 2004), Kaiser’s criterion was also considered. Kaiser (1960) recommends that the factors with eigenvalues greater than 1 should be retained for rotation (Kaiser, 1960). Although this criterion may be too strict, it is beneficial to use it when the number of items is small (<40) and sample size is greater than 205 (Gorsuch, 1983). In this study, only factors with large eigenvalues were retained ($eigenvalue > 1$).

**Factor rotation.** Factor rotation was conducted to enhance clarity in interpreting the factors (Costello & Osborne, 2005). Factor rotation maximizes the number of high item loadings and minimizes the number of low item loadings on factors to have a simple structure (Rummel, 1970). There are two types of common factor rotation techniques –
orthogonal and oblique. Orthogonal rotation assumes uncorrelated factors, and oblique rotation assumes correlated factors (Brown, 2009). If the factors are hypothesized to be independent, it is recommended to use orthogonal rotation; however, if the factors are hypothesized to be correlated, oblique rotation is preferred (Yong & Pearce, 2013). Orthogonal rotation includes varimax, quartimax, and equamax methods. Oblique rotation includes direct oblimin and promax methods. Tabachnick and Fidell (2007) discuss that “perhaps the best way to decide between orthogonal and oblique rotation is to request oblique rotation” (p. 646) because oblique rotation has better description and interpretation of relationships than orthogonal rotation (Beavers et al., 2013). Also, most of the literature recommends using oblique rotations (Costello & Osborne, 2005) and it is more appropriate for social science research (Beavers et al., 2013). Since the methods of oblique rotation are comparable, the choice generally depends on the options of the software used (Fabrigar, Wegener, MacCallum, & Strahan, 1999). In this study, promax rotation was selected. After factor extraction and rotation, the factors were named.

**Cronbach’s Coefficient Alpha**

Regarding research question 1C and research question 1D, the internal consistency reliability of the *RIRP* was examined. For Likert-scale items, Cronbach’s Coefficient Alpha is recommended (Ercan, Yazici, Sigirli, Ediz, & Kan, 2007). Cronbach’s Alpha examines the inter-relatedness of items in a measure (Tavakol & Dennick, 2011) and assesses how consistently the items on the test measure the same construct (Ary et al., 2010). The values of the Cronbach’s Alpha range from 0.00 to 1.00. A value that is closer to 1.00 indicates a higher level of reliability (Cronbach, 1951).
Therefore, Cronbach’s Coefficient Alpha was used to assess the internal consistency reliability of the RIRP.

To answer research question 2, each sub-question was considered separately. There is one dependent variable and one independent variable with two levels in each sub-question. The dependent variable was teachers’ total score on the RIRP for all these sub-questions. For research question 2A and 2B, the independent variable was teacher experience (i.e., low experience or high experience). For research question 2C and 2D, the independent variable was teachers’ level of education (i.e., bachelor’s degree or master’s degree). Also, the independent variable for research question 2E was teacher type (i.e., general education teacher or special education teacher). The following section includes the independent sample t-test procedures and the Mann-Whitney U Test procedures.

**Independent Sample t-Test Procedures**

An independent sample t-test was administered for research question 2A as the independent variable of the question has two levels (i.e., low experience or high experience). The levels of the teacher experience variable were separated according the sample distribution and size. Teachers with 15 years of experience and below were considered “low experience” and teachers with 15 years or more were considered “high experience”.

A t-test indicates whether there is a statistical difference between two means (Lomax & Hahs-Vaughn, 2013). In this case, to examine whether there was a difference between the ratings of low experienced special education teachers and high experienced
special education teachers, a t-test was considered. Also, since the samples are totally unrelated from one another for the sub-question, an independent sample of t-test was conducted.

Before administering the independent sample t-test, the non-statistical and statistical assumptions needed to be examined. Regarding the non-statistical assumptions, the dependent variable needs to be measured on a continuous variable (an interval or ratio scale) while the independent variable needs to be nominal or ordinal (Lomax & Hahs-Vaughn, 2013). Regarding the statistical assumptions, normality, independence, and homogeneity of variance were tested. The assumption of normality was detected by testing the Q-Q Plots and Shapiro-Wilk (S-W) test. For the assumption of independence, scores of participants on the dependent variable need to be independent of one another (Hinkle, Wiersma, & Jurs, 2003) meaning that one participant’s score does not depend on any other participants’ scores. Last, the assumption of homogeneity of variance was tested. The homogeneity of variance assumption requires the variance of the participants’ scores on the dependent variable to be equal. It means that scores show equal degree of variability within all levels of the independent variable (Hinkle et al., 2003). To test this assumption, Levene’s test of equality was examined.

**Mann-Whitney U Test Procedures**

The Mann-Whitney U Test is the nonparametric version of the t-test. The requirement of the both tests is to have independent samples groups. Also, they assess whether there is a difference between these groups (McKnight & Najab, 2010). However, while the Mann-Whitney U Test does not require a specific distribution (Nachar, 2008),
t-test assumes the normality (Winter & Dodou, 2010). Thus, these two tests were considered as conceptually similar and it is recommended that the Mann-Whitney $U$ Test is needed to be used when data do not meet t-test assumptions (McKnight & Najab, 2010).

The normality for each research question (i.e., 2B, 2C, 2D, and 2E) was detected by testing the Q-Q Plots and S-W test. For this study, the Mann-Whitney $U$ Test was the most appropriate statistical technique as the assumption of normality was violated for these sub-questions. For research question 2B, the Mann-Whitney $U$ Test was conducted to examine if there is a difference between the acceptability ratings of general education teachers who are in “low experience” group and those who are in “high experience” group. The critical value was calculated to determine whether there is a difference between the mean ranks of these two groups.

For research question 2C, the Mann-Whitney $U$ Test was administered to investigate whether the acceptability ratings of special education teachers on the $RIRP$ differ by their level of education (i.e., bachelor’s degree or master’s degree). The mean ranks of special education teachers with a bachelor’s degree and those with a master’s degree were examined to find out whether there is a difference on the acceptability ratings of these two groups.

Regarding to research question 2D, the acceptability ratings of general education teacher on the $RIRP$ was examined by considering their level of education. The Mann-Whitney $U$ Test was used to investigate whether there is a difference between the mean
ranks of general education teachers with bachelor’s degree and those with master’s degree.

Lastly, research question 2E was considered because the factor structure of the RIRP for special education teachers and the factor structure of the RIRP for general education teachers were found both unidimensional. Therefore, the Mann-Whitney U Test was conducted to examine whether the acceptability ratings of special education teachers on the RIRP different than general education teachers. The mean ranks of the two groups were compared to find out the difference between the groups.
CHAPTER 3
RESULTS

Overview

In this study, I examined the psychometric properties of the *Reading Intervention Rating Profile (RIRP)*, which is designed to assess teachers’ acceptability of reading interventions. Additionally, I compared ratings on the *RIRP* across different teacher characteristics. This chapter is divided into four sections: 1) descriptive statistics for the study participants; 2) results of the exploratory factor analyses; 3) findings related to the internal consistency reliability of the *RIRP*; and 4) comparison of *RIRP* ratings across various teacher characteristics.

Descriptive Statistics

In this section, I present descriptive statistics for 1) all participants; 2) special education teachers only; and 3) general education teachers only. Descriptive data included items corresponding to three categories of participant information. The first category (*demographics*) included age, gender, and race/ethnicity. The second category (*professional training and experience*) included education level (bachelor’s, master’s), teaching experience (years teaching, experience teaching struggling readers), and type of teaching license or certification (general education, special education). The third category (*instructional/organizational variables*) included grade assignment/grade(s) taught, programs/approaches participants use to teach reading, and school-wide implementation (or lack thereof) of response to intervention (RtI).
Participants

A total of 423 teachers from a Midwest state consented to participate in this study. Some participants were excluded from the final sample/data analyses for the following reasons: 1) submitted an incomplete survey \((n = 157)\); 2) indicated no experience with struggling readers or fell outside of specified grade range (grades K-5; \(n = 107\)); and/or 3) marked all Likert items “1” or “6” and were considered extreme cases in the data set \((n = 22)\). These extreme values could be considered as careless responses that were examined by looking the participants’ response pattern and time (Meade & Craig, 2012). These extreme values could be deleted from the data set (Sue & Rittens, 2007). Therefore, these cases were removed from the final analysis sample.

Thus, the final study sample included a total of 137 teachers (59 special education teachers and 78 general education teachers).

Participant demographics. The ages of all participants combined \((N = 137)\) ranged from 23 to 67 \((M = 43.8; SD = 9.97; \text{see Table 1})\), with the majority (71%) of the participants between 34 and 55 years old. Also, the majority of teachers were female (95%), while only 5% of participants were male. Last, Table 1 shows the race/ethnicity of the participants. The majority of participants were Caucasian/White (99%). Only one participant (<1%) reported multiple races as his/her ethnicity.
Table 1

**Participant Demographics (N = 137)**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 to 33</td>
<td>23</td>
<td>16</td>
</tr>
<tr>
<td>34 to 43</td>
<td>43</td>
<td>32</td>
</tr>
<tr>
<td>44 to 55</td>
<td>52</td>
<td>38</td>
</tr>
<tr>
<td>56 to 67</td>
<td>19</td>
<td>14</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Female</td>
<td>130</td>
<td>95</td>
</tr>
<tr>
<td><strong>Race/Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple Races</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Caucasian/White</td>
<td>136</td>
<td>99</td>
</tr>
</tbody>
</table>

**Professional training & experience – all participants.** Approximately 25% of participants had a bachelor’s degree, while 75% held a master’s degree (see Table 2). The proportion of participants with a master’s degree was larger than the proportion of participants with a bachelor’s degree (see Table 2). Years of teaching experience ranged from 2 to 34 ($M = 16.64; SD = 8.12$). Participants ($n = 82$) with more than 15 years of teaching experience outnumbered participants ($n = 55$) with 15 years of teaching experience or fewer. Finally, 43% of the participants have a special education certification, while 57% of the participants have a general education certification.
Table 2

*Professional Training & Experience, All Participants (N = 137)*

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Education Level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>32</td>
<td>23</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>105</td>
<td>77</td>
</tr>
<tr>
<td><strong>Teaching Experience</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 years or fewer</td>
<td>55</td>
<td>40</td>
</tr>
<tr>
<td>More than 15 years</td>
<td>82</td>
<td>60</td>
</tr>
<tr>
<td><strong>Certification</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Special Education</td>
<td>59</td>
<td>43</td>
</tr>
<tr>
<td>General Education</td>
<td>78</td>
<td>57</td>
</tr>
</tbody>
</table>

**Instructional/organizational variables – all participants.** Participants reported teaching assignment (grade or grades taught) as one of the following: K-3, 4-5, or both (two or more grades spanning K-5). The largest percentage of participants (47%) taught students in K-3rd grade, followed by 4th-5th grade (33%), and both/K-5th grade (20%).

When asked which curricula, programs, and/approaches participants used to teach reading, 57 different reading programs were identified and reported. The approaches of the programs were obtained based on the information on their websites and program examples. These programs were categorized as one of the following three reading approaches: (1) phonics based, (2) whole language, and (3) balanced, which is a combination of phonics based reading and whole language approach. Overall, 26% of the participants used a phonics based reading program, while 35% used a balanced reading program. Also, 2% of the participants used reading programs based on the whole language approach. Twenty-eight percent of the participants reported using more than
one reading program in their school (i.e., using both phonics based and balanced reading programs), while 9% reported that they did not use a specific program. Finally, 10% of the participants indicated that their school does not implement RtI in their school, while the overwhelming majority (90%) use RtI in the school (see Table 3).

Table 3

*Instructional/Organizational Variables, All Participants (N = 137)*

<table>
<thead>
<tr>
<th>Grade Level Taught</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-3</td>
<td>65</td>
<td>47</td>
</tr>
<tr>
<td>4-5</td>
<td>45</td>
<td>33</td>
</tr>
<tr>
<td>K-5</td>
<td>27</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Program/Approach</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonics-Based</td>
<td>36</td>
<td>26</td>
</tr>
<tr>
<td>Balanced</td>
<td>48</td>
<td>35</td>
</tr>
<tr>
<td>Whole Language</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Both&lt;sup&gt;a&lt;/sup&gt;</td>
<td>38</td>
<td>28</td>
</tr>
<tr>
<td>None</td>
<td>12</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>School Implementing RtI</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>123</td>
<td>90</td>
</tr>
<tr>
<td>No</td>
<td>14</td>
<td>10</td>
</tr>
</tbody>
</table>

<sup>a</sup> Both refers to teachers who use both phonics-based and balanced approaches to teach reading

**Special Education Teacher Subsample**

Several of the analyses I conducted to answer my research questions involved disaggregating participants according to whether they were special or general education teachers. Thus, I present descriptive data for the special education teacher subsample (<i>n = 59</i>), followed by the general education subsample (<i>n = 78</i>) below.
Special education teacher demographics. The age range of responding special education teachers was 23 to 67 ($M = 40.2$; $SD = 11.16$). The majority (34%) of special education teachers were between 23 and 33 years old. As indicated in Table 4, almost all of the special education teachers were female (97%), while only 3% of participants were male. Most special education teachers identified themselves as Caucasian/White (98%). One special education teacher reported *multiple races* as his/her race/ethnicity.

Table 4

*Special Education Teacher Demographics (n = 59)*

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 23 to 33</td>
<td>20</td>
<td>34</td>
</tr>
<tr>
<td>Age 34 to 43</td>
<td>16</td>
<td>27</td>
</tr>
<tr>
<td>Age 44 to 55</td>
<td>16</td>
<td>27</td>
</tr>
<tr>
<td>Age 53 to 67</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>Gender Male</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Gender Female</td>
<td>57</td>
<td>97</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple Races</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Caucasian/White</td>
<td>58</td>
<td>98</td>
</tr>
</tbody>
</table>

Professional training & experience – special education teachers. Table 5 provides information regarding special education teachers’ educational level, indicating that 29% had a bachelor’s degree. On the other hand, 71% of the special education teachers had a master’s degree. These results showed that the majority of special education teachers had a master’s degree compared to those with a bachelor’s degree (see Table 5). Total years of teaching experience ranged from 2 to 30 ($M = 12.69$; $SD = 7.93$).
Table 5 indicates that the percentage of special education teachers who have 15 years of experience and below was higher than the teachers with 15 years and above. Sixty-eight percent of special education teachers had 15 years of experience and below, and 32% had 15 years and more than 15 years of teaching experience.

Table 5

*Professional Training & Experience, Special Education Teachers (n = 59)*

<table>
<thead>
<tr>
<th>Education Level</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor’s degree</td>
<td>17</td>
<td>29</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>42</td>
<td>71</td>
</tr>
<tr>
<td>Teaching Experience</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 years or fewer</td>
<td>40</td>
<td>68</td>
</tr>
<tr>
<td>More than 15 years</td>
<td>19</td>
<td>32</td>
</tr>
</tbody>
</table>

**Instructional/organizational variables – special education teachers.** Table 6 provides information about special education teachers’ current grade level(s) taught and reading programs and/or instructional approaches that they use in their classroom. Results indicated that 48% of the special education teachers currently instruct K-3rd grade, followed by 4th-5th grade teachers (22%), and those who teach both groups (31%). Also, the majority of special education teachers (37%) implement phonics based reading programs in their classroom, 27% of them use a balanced reading program, and 32% use both phonics based and balanced reading programs in their schools. Additionally, 4% of the special education teachers do not use any program in their classroom. Finally, 10% of
the participants indicated that they do not implement RtI in the school and 90% indicated that they do implement RtI (see Table 6).

Table 6

*Instructional/Organizational Variables, Special Education Teachers (n = 59)*

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grade Level Taught</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K-3</td>
<td>28</td>
<td>48</td>
</tr>
<tr>
<td>4-5</td>
<td>13</td>
<td>22</td>
</tr>
<tr>
<td>K-5</td>
<td>18</td>
<td>31</td>
</tr>
<tr>
<td><strong>Program/Approach</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonics Based</td>
<td>22</td>
<td>37</td>
</tr>
<tr>
<td>Balanced</td>
<td>16</td>
<td>27</td>
</tr>
<tr>
<td>Both*</td>
<td>19</td>
<td>32</td>
</tr>
<tr>
<td>None</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td><strong>School Implementing RtI</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>53</td>
<td>90</td>
</tr>
<tr>
<td>No</td>
<td>6</td>
<td>10</td>
</tr>
</tbody>
</table>

*Both refers to teachers who use both phonics based and balanced reading approach.*

**General Education Teacher Subsample**

A total of 78 general education teachers were included in this study. Descriptive statistics for the general education teacher subsample are presented next.

**General education teacher demographics.** The general education subsample contained 78 teachers. Table 7 shows that their ages ranged from 26 to 62 (M = 46.6; SD = 7.98). The results showed that 54% of the general education teachers’ ages were between 40 and 49 years. Regarding the participants’ gender, 94% of teachers were female, while 6% were male (see Table 7). Additionally, all general education teachers
reported Caucasian/White as their race/ethnicity \( (n=78) \), with no other races/ethnicities reported.

Table 7

*General Education Teacher Demographics \( (n = 78) \)*

<table>
<thead>
<tr>
<th></th>
<th>( n )</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 to 39</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>40 to 49</td>
<td>42</td>
<td>54</td>
</tr>
<tr>
<td>50 to 62</td>
<td>25</td>
<td>32</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Female</td>
<td>73</td>
<td>94</td>
</tr>
<tr>
<td><strong>Race/Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian/White</td>
<td>78</td>
<td>100</td>
</tr>
</tbody>
</table>

**Professional training & experience – general education teachers.** In the second section of the survey, 19\% of the participants indicated that they had a bachelor’s degree, while 81\% held a master’s degree (see Table 8). Also, for general education teachers’, teaching experience in total number of years ranged from 3 to 34 \((M =19.62; SD = 6.94)\). Last, Table 8 indicates that the participants with 15 years or more of experience outnumbered those with 15 years of experience or less.
Table 8

*Professional Training & Experience, General Education Teachers (n = 78)*

<table>
<thead>
<tr>
<th>Education Level</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor’s degree</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>63</td>
<td>81</td>
</tr>
<tr>
<td>Teaching Experience</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 years or fewer</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>More than 15 years</td>
<td>63</td>
<td>81</td>
</tr>
</tbody>
</table>

**Instructional/organizational variables** – **general education teachers.** Just under half (47%) of the general education teachers were assigned to teach students in K-3rd grade. Those who teach 4th-5th grade comprise 32% of the sample, and teachers instructing both groups total 12% (see Table 9). The majority of the general education teachers (41%) use a balanced reading program, 18% use a phonics based reading program, and 24% use both phonics based and balanced reading programs. Moreover, 4% use a whole language approach, and 13% do not use any program in their classroom (see Table 9). Regarding the use of RtI in the schools, 10% of general education teachers reported that their school does not implement RtI, and 90% indicated that their school does (see Table 9).
Table 9

*Instructional/Organizational Variables, General Education Teachers (n = 78)*

<table>
<thead>
<tr>
<th>Demographic Information</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade Level Taught</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K-3</td>
<td>37</td>
<td>47</td>
</tr>
<tr>
<td>4-5</td>
<td>32</td>
<td>41</td>
</tr>
<tr>
<td>K-5</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>Reading Approaches</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonics Based</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td>Balanced</td>
<td>32</td>
<td>41</td>
</tr>
<tr>
<td>Whole</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Both^a</td>
<td>19</td>
<td>24</td>
</tr>
<tr>
<td>None</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>Implementing RtI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>70</td>
<td>90</td>
</tr>
<tr>
<td>No</td>
<td>8</td>
<td>10</td>
</tr>
</tbody>
</table>

*Both refers to teachers who use both phonics based and balanced reading approach.*

Results from Data Analyses

In this section, I first present results from the exploratory factor analysis. Next, I detailed the results regarding the internal consistency reliability of the *RIRP*. Last, I describe results from comparisons of the *RIRP* ratings by various teacher characteristics.

Exploratory Factor Analyses Results

The results of statistical assumptions, factor extraction, and rotation are explained as follows.

*Tests of statistical assumptions.* Prior to investigating the factor structure of the *RIRP*, I tested the following assumptions to ensure that the data were suitable for conducting EFA. I first tested for linear relations among variables by examining
scatterplots. Careful inspection of the scatterplots suggested that the assumption of linearity was met as the scatterplots followed a linear pattern. Also, I tested for multicollinearity within data set. The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy and Bartlett’s test of sphericity were used to determine the factorability. The KMO index ranges from 0 to 1, with a magnitude equal to or higher than .50 considered as acceptable. The results indicated that sample size was adequate for the 15-item scale, and the correlation matrix was considered factorable. The results of the KMO and Bartlett’s test of sphericity are included below for each subgroup.

**Factor extraction and rotation.** For the special education teacher subgroup, I used Principal Axis Factoring (PAF) with Promax rotation to examine the underlying factor structure of the 15-item scale, as recommended by Costello and Osborne (2005). The KMO test result was found to be acceptable (Kaiser, 1974) with a value of .928, and Bartlett’s test of sphericity was significant ($\chi^2 = 1035.763, df = 105, p < .001$). The results indicated the following: (1) the correlation matrix was not an identity matrix and the variables were factorable, and (2) the sample size was adequate for the 15-item scale used. The EFA results produced one general acceptability factor accounting for 72.0% of the variance (see Table 10).
Table 10

*Factor Matrix from the RIRP for Special Education Teachers (N=59)*

<table>
<thead>
<tr>
<th>Items</th>
<th>Factor 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall, the intervention would be beneficial for the child.</td>
<td>.966</td>
</tr>
<tr>
<td>The intervention is reasonable for the reading problem described.</td>
<td>.949</td>
</tr>
<tr>
<td>The intervention was a fair way to handle the child’s reading problem.</td>
<td>.923</td>
</tr>
<tr>
<td>This intervention was a good way to handle the child’s reading problem.</td>
<td>.921</td>
</tr>
<tr>
<td>I would suggest the use of this intervention to other teachers.</td>
<td>.917</td>
</tr>
<tr>
<td>The intervention is consistent with those I have used in classroom settings.</td>
<td>.875</td>
</tr>
<tr>
<td>This would be an acceptable intervention for the child’s reading problem.</td>
<td>.870</td>
</tr>
<tr>
<td>The intervention would increase the child’s reading skills and achievement.</td>
<td>.843</td>
</tr>
<tr>
<td>I would be willing to use this in the classroom setting.</td>
<td>.840</td>
</tr>
<tr>
<td>I like the procedures used in the intervention.</td>
<td>.837</td>
</tr>
<tr>
<td>Most teachers would find this intervention suitable for the reading problem described.</td>
<td>.798</td>
</tr>
<tr>
<td>The child’s reading problem is severe enough to warrant use of this intervention.</td>
<td>.741</td>
</tr>
<tr>
<td>Most teachers would find this intervention appropriate for reading problems in addition to the one described.</td>
<td>.705</td>
</tr>
<tr>
<td>The intervention would be an appropriate intervention for children at varying levels of reading skills.</td>
<td>.667</td>
</tr>
<tr>
<td>The intervention would not result in negative side-effects for the child.</td>
<td>.637</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Axis Factoring.

The same procedures were used for the general education teacher subgroup. PAF extraction with Promax rotation was selected to investigate the underlying factor structure of the scale. The KMO test result was acceptable, with value was .944, and Bartlett’s Test of Sphericity was significant ($\chi^2 = 1467.084, df = 105, p < .001$). The
results indicated that the correlation matrix was not an identity matrix and the variables were factorable. Additionally, results showed that the sample size was adequate for analysis of the 15-item scale. According to the EFA results, one general acceptability factor emerged accounting for 74.0% of the variance (see Table 11).
Table 11

*Factor Matrix from the RIRP for General Education Teachers (N=78)*

<table>
<thead>
<tr>
<th>Items</th>
<th>Factor 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall, the intervention would be beneficial for the child.</td>
<td>.944</td>
</tr>
<tr>
<td>The intervention is reasonable for the reading problem described.</td>
<td>.941</td>
</tr>
<tr>
<td>I would suggest the use of this intervention to other teachers.</td>
<td>.937</td>
</tr>
<tr>
<td>The intervention was a fair way to handle the child’s reading problem.</td>
<td>.925</td>
</tr>
<tr>
<td>I would be willing to use this in the classroom setting.</td>
<td>.921</td>
</tr>
<tr>
<td>This intervention was a good way to handle the child’s reading problem.</td>
<td>.903</td>
</tr>
<tr>
<td>This would be an acceptable intervention for the child’s reading problem.</td>
<td>.899</td>
</tr>
<tr>
<td>The intervention would increase the child’s reading skills and achievement.</td>
<td>.898</td>
</tr>
<tr>
<td>I like the procedures used in the intervention.</td>
<td>.854</td>
</tr>
<tr>
<td>The intervention would be an appropriate intervention for children at varying levels of reading skills.</td>
<td>.823</td>
</tr>
<tr>
<td>The intervention is consistent with those I have used in classroom settings.</td>
<td>.804</td>
</tr>
<tr>
<td>Most teachers would find this intervention suitable for the reading problem described.</td>
<td>.776</td>
</tr>
<tr>
<td>The child’s reading problem is severe enough to warrant use of this intervention.</td>
<td>.770</td>
</tr>
<tr>
<td>The intervention would not result in negative side-effects for the child.</td>
<td>.669</td>
</tr>
<tr>
<td>Most teachers would find this intervention appropriate for reading problems in addition to the one described.</td>
<td>.622</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Axis Factoring.

**Cronbach’s Coefficient Alpha Results**

The internal consistency reliability of the *RIRP* was examined. For special education subgroup, Cronbach’s Coefficient Alpha indicated a high level of internal consistency (α = .971). In addition, the item means range was 3.58 to 4.17 (see Table
12), and inter-item correlations were moderate and statistically significant between all pairs of items ($p < .01$).

Table 12

*Descriptive Statistics for the RIRP for Special Education Teachers ($N = 59$)*

<table>
<thead>
<tr>
<th>Items (N = 15)</th>
<th>Min</th>
<th>Max</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>This would be an acceptable intervention for the child’s reading problem.</td>
<td>1</td>
<td>6</td>
<td>4.03</td>
<td>1.231</td>
</tr>
<tr>
<td>Most teachers would find this intervention appropriate for reading problems in addition to the one described.</td>
<td>1</td>
<td>6</td>
<td>4.02</td>
<td>1.106</td>
</tr>
<tr>
<td>The intervention would increase the child’s reading skills and achievement.</td>
<td>1</td>
<td>6</td>
<td>3.97</td>
<td>1.159</td>
</tr>
<tr>
<td>I would suggest the use of this intervention to other teachers.</td>
<td>1</td>
<td>6</td>
<td>3.78</td>
<td>1.301</td>
</tr>
<tr>
<td>The child’s reading problem is severe enough to warrant use of this intervention.</td>
<td>1</td>
<td>6</td>
<td>4.12</td>
<td>1.327</td>
</tr>
<tr>
<td>Most teachers would find this intervention suitable for the reading problem described.</td>
<td>1</td>
<td>6</td>
<td>4.17</td>
<td>1.234</td>
</tr>
<tr>
<td>I would be willing to use this in the classroom setting.</td>
<td>1</td>
<td>6</td>
<td>3.92</td>
<td>1.381</td>
</tr>
<tr>
<td>The intervention would not result in negative side-effects for the child.</td>
<td>1</td>
<td>6</td>
<td>4.14</td>
<td>1.279</td>
</tr>
<tr>
<td>The intervention would be an appropriate intervention for children at varying levels of reading skills.</td>
<td>1</td>
<td>6</td>
<td>3.64</td>
<td>1.448</td>
</tr>
<tr>
<td>The intervention is consistent with those I have used in classroom settings.</td>
<td>1</td>
<td>6</td>
<td>3.88</td>
<td>1.391</td>
</tr>
<tr>
<td>The intervention was a fair way to handle the child’s reading problem.</td>
<td>1</td>
<td>6</td>
<td>3.73</td>
<td>1.311</td>
</tr>
<tr>
<td>The intervention is reasonable for the reading problem described.</td>
<td>1</td>
<td>6</td>
<td>3.76</td>
<td>1.318</td>
</tr>
<tr>
<td>I like the procedures used in the intervention.</td>
<td>1</td>
<td>6</td>
<td>3.68</td>
<td>1.370</td>
</tr>
<tr>
<td>This intervention was a good way to handle the child’s reading problem.</td>
<td>1</td>
<td>6</td>
<td>3.58</td>
<td>1.342</td>
</tr>
<tr>
<td>Overall, the intervention would be beneficial for the child.</td>
<td>1</td>
<td>6</td>
<td>3.88</td>
<td>1.274</td>
</tr>
</tbody>
</table>

*Note.* Strongly Disagree = 1, Disagree = 2, Slightly Disagree = 3, Slightly Agree = 4, Agree = 5, and Strongly Agree = 6.
The internal consistency reliability of the RIRP was also examined for the general education teacher RIRP scores. Cronbach’s Coefficient Alpha indicated a high level of internal consistency ($\alpha = .974$). Lastly, item means ranged from 3.05 to 3.85 (see Table 13), and inter-item correlations were moderate and statistically significant ($p < .01$).
Table 13

*Descriptive Statistics for the RIRP for General Education Teachers (N = 78)*

<table>
<thead>
<tr>
<th>Items (N = 15)</th>
<th>Min</th>
<th>Max</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>This would be an acceptable intervention for the child’s reading problem.</td>
<td>1</td>
<td>6</td>
<td>3.40</td>
<td>1.390</td>
</tr>
<tr>
<td>Most teachers would find this intervention appropriate for reading problems in addition to the one described.</td>
<td>1</td>
<td>6</td>
<td>3.79</td>
<td>1.323</td>
</tr>
<tr>
<td>The intervention would increase the child’s reading skills and achievement.</td>
<td>1</td>
<td>6</td>
<td>3.44</td>
<td>1.244</td>
</tr>
<tr>
<td>I would suggest the use of this intervention to other teachers.</td>
<td>1</td>
<td>6</td>
<td>3.22</td>
<td>1.392</td>
</tr>
<tr>
<td>The child’s reading problem is severe enough to warrant use of this intervention.</td>
<td>1</td>
<td>6</td>
<td>3.85</td>
<td>1.495</td>
</tr>
<tr>
<td>Most teachers would find this intervention suitable for the reading problem described.</td>
<td>1</td>
<td>6</td>
<td>3.82</td>
<td>1.326</td>
</tr>
<tr>
<td>I would be willing to use this in the classroom setting.</td>
<td>1</td>
<td>6</td>
<td>3.40</td>
<td>1.462</td>
</tr>
<tr>
<td>The intervention would not result in negative side-effects for the child.</td>
<td>1</td>
<td>6</td>
<td>3.50</td>
<td>1.475</td>
</tr>
<tr>
<td>The intervention would be an appropriate intervention for children at varying levels of reading skills.</td>
<td>1</td>
<td>6</td>
<td>3.45</td>
<td>1.483</td>
</tr>
<tr>
<td>The intervention is consistent with those I have used in classroom settings.</td>
<td>1</td>
<td>6</td>
<td>3.44</td>
<td>1.401</td>
</tr>
<tr>
<td>The intervention was a fair way to handle the child’s reading problem.</td>
<td>1</td>
<td>6</td>
<td>3.22</td>
<td>1.364</td>
</tr>
<tr>
<td>The intervention is reasonable for the reading problem described.</td>
<td>1</td>
<td>6</td>
<td>3.40</td>
<td>1.361</td>
</tr>
<tr>
<td>I like the procedures used in the intervention.</td>
<td>1</td>
<td>6</td>
<td>3.17</td>
<td>1.253</td>
</tr>
<tr>
<td>This intervention was a good way to handle the child’s reading problem.</td>
<td>1</td>
<td>6</td>
<td>3.05</td>
<td>1.357</td>
</tr>
<tr>
<td>Overall, the intervention would be beneficial for the child.</td>
<td>1</td>
<td>6</td>
<td>3.32</td>
<td>1.344</td>
</tr>
</tbody>
</table>

*Note.* Strongly Disagree = 1, Disagree = 2, Slightly Disagree = 3, Slightly Agree = 4, Agree = 5, and Strongly Agree = 6.
Independent Samples t-Test Results

An independent sample t-test was used to investigate the difference between “low experience” special education teachers and “high experience” special education teachers on the acceptability ratings from the RIRP. First I tested underlying statistical assumptions were tested. Results of Q-Q Plots and the Shapiro-Wilk (S-W) tests revealed that the assumption of normality was met ($p > .05$). For the assumption of independence, participants’ scores on the RIRP were independent of one another, in that one participant’s score is not related to any other participants’ scores. In addition, I tested the assumption of homogeneity of variance using Levene’s Test of Equality of Variances, which indicated that the assumption of homogeneity of variance was violated ($p = .001$).

To create the independent variable groups, I also examined the sample distribution and size. Teachers with 15 years of experience and below were considered “low experience” ($n = 40$) and teachers with 15 years or more were considered “high experience” ($n = 19$). The results indicated a significant difference on the ratings of the RIRP between “low experience” special education teachers ($M = 4.27, SD = .76$) and “high experience” special education teachers ($M = 3.07, SD = 1.26; t [24.456] = 3.8, p = .001$). These results suggest that the special education teachers with fewer years of experience rated the RIRP items as more acceptable when compared to those with more years of experience.

Additionally, I calculate the effect size ($d$) of the difference between the ratings of “low experience” special education teachers and “high experience” special education teachers by considering Cohen’s (1988) guidelines recommending that if $d = .20$, the
effect size is considered “small”; if $d = .50$, the effect size is considered “medium”; and if $d = .80$, the effect size is considered “large”. According to these guidelines, the magnitude of the difference between two groups found in this analysis was considered “large” as $d = 1.15$.

**The Mann-Whitney $U$ Test Results**

The non-parametric Mann-Whitney $U$ Test was used to investigate differences between various teacher demographic groups on the acceptability ratings from the *RIRP*. The Mann-Whitney $U$ Test is used if the assumption of normality is violated.

First, I examined the difference between general education teachers who are in “low experience” group ($n = 15$) and those who are in “high experience” group ($n = 63$). Normality was evaluated by testing Q-Q Plots and the S-W test. The results indicated that normality was violated ($p < .05$). The small sample size used for this study is most probably a contributor of experiencing non-normality. The Mann-Whitney $U$ Test results showed that the difference between the two groups was statistically significant ($Z = -2.239, p = .025$). Therefore, the two groups of teachers did not have similar average rank scores on their *RIRP* ratings. Additionally, Table 14 shows that teachers who are in “low experience” group had higher ranks compared to teachers who are in “high experience” group. Thus, general education teachers who have fewer years of experience rated the *RIRP* items as more acceptable than those who have more years of experience (see Table 14).
Table 14

Mean Ranks and Sum of Ranks for the Ratings of General Education Teachers on the RIRP - Experience

<table>
<thead>
<tr>
<th>Ranks$^a$</th>
<th>Experience</th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>RIRP</td>
<td>15 and below</td>
<td>15</td>
<td>51.27</td>
<td>769.00</td>
</tr>
<tr>
<td></td>
<td>Above 15</td>
<td>63</td>
<td>36.70</td>
<td>2312.00</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>78</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Last, the effect size for the Mann-Whitney U test can be calculated according to the following formula (Fritz, Morris, & Richler, 2011) where $r$ is the effect size, $Z$ is the $Z$ score, and $N$ is the total sample size:

$$r = \frac{Z}{\sqrt{N}}$$

Based on Cohen’s guidelines for $r$, .50 is considered “large” effect, .30 is considered “medium” effect, and .10 is considered “small” effect (Coolican, 2009). The calculation showed that the effect size of the difference between the ratings of “low experience” general education teachers and “high experience” general education teachers for the analysis was small as $r = .25$. The results suggested that even thought the findings were significant; the magnitude of the difference between two groups was small.

Second, the Mann-Whitney U Test was conducted to examine whether the acceptability ratings of special education teachers on the RIRP differ by their level of education (i.e., bachelor’s degree and master’s degree). Q-Q Plots and the S-W test indicated that data were not normally distributed ($p < .05$). The small sample size used for this study and having unequal number of participants regarding the certification type are most probably the contributors of experiencing non-normality.
The results of the Mann-Whitney $U$ Test showed that there was no significant difference on the ratings of the $RIRP$ items between special education teachers with a bachelor’s degree ($n = 17$) and those with a master’s degree ($n = 42$; $Z = -.243, p = .808$). The results indicated that the two groups of teachers had similarly ranked ratings on the $RIRP$. Also, Table 15 shows that the difference between the mean ranks of the two groups is small. These results suggest that the acceptability ratings of the special education teachers with a bachelor’s degree were not different on the $RIRP$ compared to those who have a master’s degree (see Table 15).

Table 15

<table>
<thead>
<tr>
<th>Ranks</th>
<th>Education</th>
<th>$n$</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RIRP$</td>
<td>Bachelor’s</td>
<td>17</td>
<td>29.15</td>
<td>495.50</td>
</tr>
<tr>
<td></td>
<td>Master’s</td>
<td>42</td>
<td>30.35</td>
<td>1274.50</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>59</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Third, the Mann-Whitney $U$ Test was used to examine whether the acceptability ratings of general education teachers on the $RIRP$ differed by their respective level of education (i.e., bachelor’s degree and master’s degree). The Q-Q Plots and the S-W test were conducted to determine if the data were normally distributed. The results indicated that normality was violated ($p < .05$) which is likely a result of using small sample size and unequal number of participants from each subgroup (i.e., number of participants with bachelor’s degree and those with master’s degree). The results of the Mann-Whitney $U$ Test showed that there was no significant difference on the ratings of the $RIRP$ between
general education teachers with a bachelor’s degree and those with a master’s degree ($Z = -0.076, p = .939$). Table 16 shows there is no significant difference between the mean ranks of the two groups. These results suggest that the acceptability ratings of general education teachers with a bachelor’s degree and those with a master’s degree are statistically equivalent (See Table 16).

Table 16

<table>
<thead>
<tr>
<th>Ranks</th>
<th>Education</th>
<th>n</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bachelor’s</td>
<td>15</td>
<td>39.10</td>
<td>586.50</td>
</tr>
<tr>
<td>RIRP</td>
<td>Master’s</td>
<td>63</td>
<td>39.60</td>
<td>2494.50</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>78</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Since the factor structure of the $RIRP$ for special education teachers and for general education teachers were both unidimensional, the differences between the acceptability ratings of special education teachers and general education teachers on the $RIRP$ were examined. First, I consulted Q-Q Plots and the S-W test to determine whether the data were normally distributed. The results indicated that the data were not normally distributed ($p < .05$); possibly due to having small sample size and unequal distribution of number of participants form each subgroup (i.e., the number of special education teachers and general education teachers). Therefore, the Mann-Whitney $U$ Test was conducted to investigate group differences. The Mann-Whitney $U$ Test results showed that there was a significant difference on the ratings of the $RIRP$ between general education teachers and special education teachers ($Z = -2.236, p = .025$). In comparing the mean ranks of the
two groups, the results showed that the special education teachers had higher acceptability ratings on the *RIRP* than the general education teachers (see Table 17).

Last, based on the formula noted above, the effect size \((r)\) for the analysis was found .19 which is considered a small effect meaning that the size of the difference between special and general education teachers was small.

Table 17

*Mean Ranks and Sum of Ranks for the Ratings of Special Education Teachers and General Education Teachers on the RIRP*

<table>
<thead>
<tr>
<th>Ranks</th>
<th>Groups</th>
<th>n</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Special Education Teachers</td>
<td>59</td>
<td>77.71</td>
<td>4585.00</td>
</tr>
<tr>
<td>RIRP</td>
<td>General Education Teachers</td>
<td>78</td>
<td>62.41</td>
<td>4868.00</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>137</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 4
DISCUSSION

Overview

This chapter has three sections. In the first section, I briefly review the purpose of
the study and the research questions. In the second section, I summarize the findings and
relate them to the research questions and previous research. The final section concludes
with the study’s limitations, the implications of the results, and recommendations for
future research.

Review of Statement of Problem and Study Purpose

Reading difficulties play an important role in students’ overall academic
achievement (Lonigan, Purpura, Wilson, Walker, & Clancy-Menchetti, 2013). However,
many teachers do not use effective strategies for students who have reading problems
(Ness, 2016). While there are many reasons for teachers’ disuse of well-supported
reading practices (e.g., research-to-practice gap), treatment acceptability of these practices
is one of the most critical (Reimers et al., 1992). Treatment acceptability comprises
subjective evaluations of treatment procedures regarding their appropriateness, fairness,
and reasonableness to a given situation (Kazdin, 1981). To empirically investigate the
acceptability of interventions and factors that influence treatment acceptability, treatment
acceptability measures have been developed for both clinical and school-based
interventions.
To date, the primary focus of treatment acceptability research has been investigating the treatment acceptability of behavioral interventions specifically; however, a limited number of studies have investigated the treatment acceptability of academic interventions, and in particular reading interventions (McCurdy & Shapiro, 1992; Mautone et al., 2009; Shapiro & Goldberg, 1990; Turco & Elliott, 1990). While the need to investigate the treatment acceptability of reading interventions for struggling readers is clear (Denton et al., 2003; Haager, Klingner, & Vaughn, 2007; NPR, 2000), no validated measures have been developed to assess the treatment acceptability of such interventions. Therefore, the first purpose of this study was to develop a new instrument called the Reading Intervention Rating Profile (RIRP) which was specifically designed to assess the acceptability of reading intervention procedures. To achieve this first purpose, I examined the psychometric properties of the RIRP (i.e., factor structure for special education teachers and general education teachers, the internal consistency reliability for special education teachers and general education teachers). The second major purpose was to examine teacher factors that influence acceptability ratings of reading interventions, as measured by the RIRP. Specifically, I examined whether the acceptability ratings of reading interventions differ by teacher type (special versus general education), teaching experience, and level of education.
Discussion of the Study Findings

In this section, I provided a detailed discussion about the results of exploratory factory analyses and reliability of the RIRP and relationships between teacher characteristics and RIRP ratings.

Exploratory Factory Analyses and Reliability of the RIRP

Exploratory factor analysis (EFA) was used to explore the psychometric properties of the RIRP. I examined the factor structure and the internal consistency reliability of the RIRP for special education teachers and general education teachers.

For special education teachers, the findings suggest that special education teachers’ judgments of acceptability for reading interventions on the 15-item scale are comprised of one general acceptability factor. Specifically, the EFA results showed that the RIRP has one strong unitary factor that explains the largest proportion of the variance (72%). Moreover, for general education teachers, the EFA results also produced one general acceptability factor accounting for 74.0% of the variance meaning that one unitary acceptability factor accounted for the largest proportion of the variance. These findings are consistent with the results of various studies investigating the factor structure of treatment acceptability scales, as many studies have found one unitary acceptability factor for their scales (e.g., Kazdin, 1980; Martens et al., 1985; Tarnowski & Simonian, 1992; Witt & Elliott, 1985). On the other hand, other studies that investigated the factor structure of treatment acceptability scales found more than one factor (e.g., Kratochwill & VanSomeren; 1984; Witt & Martens, 1983). For example, Witt and Martens (1983) developed the Intervention Rating Profile (IRP) and examined the acceptability of school
based behavioral interventions. Their results indicated that the IRP has one primary acceptability factor accounting for 41% of the variance and several secondary factors i.e., risk for the target child (9% of the variance), time for teachers to implement the treatment (9% of the variance), effects on other children (7% of the variance), and skills required for teachers to use the treatment (7% of the variance). Similarly, Kratochwill and VanSomeren (1984) developed the Assessment Rating Profile (ARP) to examine the acceptability of assessment measures and methods. Like with the study of Witt and Martens (1983), the authors found the ARP to be multifactorial as it had loadings on overall acceptability factor accounted for 54.9% of variance, intrusiveness factor accounted for 10.2% of variance and danger or high risk factor accounted for 6% of variance. Both scales required refinements due to weaknesses related to having poor loadings on the secondary factors. A possible explanation for the weak loadings may be using a principal component analysis as a factor rotation method. Reliance on this rotation method may allow the scales to remain flawed (Eckert et al., 1999). Furthermore, the construct validity of the measures was seen as controversial insofar as the problematic factor loadings could be construed as a result of poor item wording and deficient content (Eckert et al., 1999). Therefore, for the current study, it could be said that finding one unitary acceptability factor with sufficient psychometric properties could be considered as confirmation of the predictability of the RIRP and initial evidence for the construct validity of the RIRP in the context of measuring the acceptability of reading interventions.

The internal consistency reliability of the RIRP for special education teachers and
general education teachers was also investigated by calculating Cronbach’s Coefficient Alpha. For special education teachers, the results suggested that the internal consistency reliability of the RIRP was strong ($\alpha = .971$) meaning that the scale produced a sufficient psychometric property. Similarly, for general education teachers, the strong internal consistency reliability of the RIRP ($\alpha = .974$) suggests that the measure has an acceptable psychometric property. These results are also consistent with previous studies in that, for all of the studies that found one general acceptability factor, the measures also exhibited strong internal consistency reliabilities (e.g. Kazdin, 1980; Martens et al., 1985; Tarnowski & Simonian, 1992; Witt & Elliott, 1985). A possible reason for the strong internal consistency reliabilities might be related to the influence of having one unitary factor, which can increase the internal consistency of the measures (Eckert, Hier, Hamsho, & Malandrino, 2017). Thus, this study revealed that ratings of special and general education teachers for the acceptability of reading interventions on the RIRP produced one unitary acceptability factor with sufficient psychometric properties.

**Relationships Between Teacher Characteristics and RIRP Ratings**

Regarding the second purpose of the study, I also investigated teachers’ acceptability ratings of reading interventions as they relate to various teacher characteristics (years of teaching experience, level of education, general versus special education role).

**Years of teaching experience.** I examined whether the acceptability ratings of teachers on the RIRP differ based on their “years of teaching experience”. For special education teachers, the results indicated that the teachers who have fewer years (< 15) of
experience rated the RIRP items as more acceptable \((M = 4.27, SD = .76)\) than those who have more years \((15+)\) of experience \((M = 3.07, SD = 1.26; t [24.456] = 3.8, p = .001)\). Also, the effect size \((d = 1.15)\) was large according to Cohen’s guidelines. Likewise, for general education teachers, the results showed that teachers who have fewer years \(< 15\) of experience rated the RIRP items as being more acceptable than did those teachers who have more years \((15+)\) of experience, and the difference between the two groups was statistically significant \((Z = -2.239, p = .025)\).

These results are similar to the findings of prior studies that indicate a negative relationship between years of teaching experience and acceptability of interventions. For example, the study by Witt and his colleagues (1984), and that of Witt and Robbins (1985) found similar results as teachers who were less experienced were more likely than highly experienced teachers to accept recommended treatments. The reason for the negative relationship between years of teaching experience and acceptability of interventions might be related to innovations within teacher preparation programs and reading practices. Across the last four decades, the attention to the amount and quality of training related to reading intervention has increased (Hoffman et al., 2005). Various national reform reports (e.g., A Nation at Risk, 1983; A Nation Prepared, 1986; What Matters Most: Teaching and America’s Future, 1996) have focused on teacher preparation programs since the mid-1980s (Brownell, Ross, Colón, & McCallum, 2005). Also, the federal law, No Child Left Behind (NCLB, 2002) Act requires implementing effective reading practices and having qualified teachers in classrooms. Therefore, the attention has been targeted to the instructional innovations in reading and teacher
preparation. For example, a survey conducted by the Commission of the International Reading Association (IRA, 1999) showed an increased focus on reading preparation within teacher preparation programs, i.e., average number of course hours in reading per semester, availability of undergraduate reading specialization, emphases on effective reading approaches, and learning how to implement reading practices for diverse learners (Hoffman et al., 2005). For the current study, then, a possible explanation of the negative relationship might be that “low experienced” teachers might have more positive attitudes toward recommended practices as they were more immersed in these practices during their preservice education.

**Level of education.** I examined whether the acceptability ratings of teachers on the RIRP differ based on their “level of education” (i.e., bachelor’s degree or master’s degree). For special education teachers, the current study revealed that the acceptability ratings of special education teachers with a bachelor’s degree were not different when compared to those who have a master’s degree ($Z = -.243, p = .808$). For general education teachers, similar results were found. The findings suggest that the acceptability ratings of general education teachers with a bachelor’s degree were similar to those with a master’s degree ($Z = -.076, p = .939$). These results are inconsistent with various previous studies indicating that teachers’ training influences their acceptability ratings (e.g. Stormont et al., 2011; Vereb & DiPerna, 2004) and use of effective practices (e.g. Kutash et al., 2009). A possible explanation of these inconsistent results might be related to differences in interventions used in the studies. For example, while the previous studies focused on teachers’ agreements of evidence-based practices for behavior
management (Stormont et al., 2011) and teachers’ acceptability ratings of different behavioral treatment approaches (Vereb & DiPerna, 2004), the current study investigated teachers’ acceptability ratings based on a reading intervention. Additionally, the instruments used in the studies were also different. In the study of Stormont and her colleagues (2011), a 17-item survey that measures teachers’ agreement ratings for evidence-based and non-evidence based practices was administered to 328 general and 35 special education teachers. Furthermore, in the study of Vereb and DiPerna (2004), a 50-item Knowledge of ADHD Rating Evaluation (KARE) survey was administered to 47 elementary school teachers. Both measures were not directly related to assess teachers’ treatment acceptability. Even though the KARE has two domains which measure acceptability (i.e., medication and behavior management), those domains make very small portion of the KARE (10 questions) and do not assess the acceptability of reading interventions. Therefore, it could be said that use of different type of interventions and different type of measurement scales might be a reason of the inconsistent results.

Another possible explanation might be related to the “use of RtI” in schools. The descriptive results from special education teachers’ sample and general education teacher’s sample indicated that the vast majority of teachers use response to intervention (RtI) in their schools. The use of effective practices is one of the critical strengths of RtI (Barnes & Harlacher, 2008) as it provides teachers with a systematic way of using research-based instruction for struggling readers (WWC, 2009). Also, RtI resources are provided to teachers for their professional development in many states (e.g. online training tools, statewide training by the state department of education, state resource
centers) along with receiving support by many stakeholders (Berkeley, Bender, Peaster, & Saunders, 2009). Since the focus of RtI includes high-quality instruction, data-based decisions, and progress monitoring, its implementation in schools could help teachers increase their understanding of effective reading practices and could change the views of teachers toward recommended practices in a positive way. As a result, the use of RtI could diminish the differences between teachers regarding the acceptability of effective reading interventions. Therefore, knowledge of effective interventions through RtI might be the reason for similar acceptability ratings on the RIRP among teachers with different levels of education.

**Special education versus general education.** Finally, the factor structure of the RIRP for special and general education teachers were found both unidimensional; therefore, the acceptability ratings of the special and general education teachers on the RIRP were examined. The ratings on the RIRP showed that special education teachers had higher acceptability scores than the general education teachers ($Z = -2.236, p = .025$). This is important in light of earlier research suggesting that special education teachers were more likely than general education teachers to accept recommended practices. For example, Easton and Erchul (2011) investigated teachers’ acceptability of “treatment plan implementation (TPI) monitoring and feedback methods” (p. 56) by using the Monitoring and Feedback Acceptability Survey (MFAS). In the study, researchers examined how the acceptability ratings of special education teachers and general education teachers differed on the MFAS regarding the TPI monitoring and feedback methods (Easton & Erchul, 2011). The researchers found that special education
teachers had higher acceptability ratings on the MFAs than did the general education teachers. Moreover, Maccini and Gagnon (2006) conducted a survey to investigate whether special education teachers and general education teachers use specific instructional adaptations for students with learning disabilities and/or emotional/behavioral disorders. The researchers reported that special education teachers were more likely to use recommended instructional strategies (i.e., peer tutoring) and assessment accommodations than general education teachers. Even though Maccini and Gagnon (2006) did not assess teachers’ acceptability ratings, their results support the idea that special education teachers are more likely to accept and implement recommended practices than general education teachers.

While the focus of previous studies was on perceived knowledge of special and general education teachers for instructional adaptations (Maccini & Gagnon, 2006), as well as the acceptability of monitoring and feedback methods (Easton & Erchul, 2011), the current study contributes to the literature by making evident the difference between the acceptability ratings of two groups of teachers regarding reading interventions. The observed effect size for the difference between general and special education teacher ratings was small; however, possibly due to the fact that general educators have increased their focus on effective reading instruction over time as the implementation of evidence-based practices in schools is mandated by federal law (Individuals with Disabilities Education Improvement Act [IDEA], 2004; NCLB, 2002). Therefore, use of RtI and other initiatives might have contributed to improvements in teachers’ reading instruction, as well as student outcomes.
Another reason for the discrepant acceptability ratings for general and special education teachers might be related to the reading programs that teachers use in their respective classrooms. The descriptive results showed that special and general education teachers are implementing different reading programs. While the majority of special education teachers use phonics-based reading programs, most of the general education teachers implement balanced reading programs. As other researchers have previously suggested, the *theoretical orientation* of a treatment is a factor that influences teachers’ acceptability ratings (Kazdin & Cole, 1981; Witt, 1986). Therefore, implementation of reading programs that have different theoretical orientations might be the reason for finding different acceptability ratings on the *RIRP*. Past research has also supported the idea that teachers’ familiarity with an intervention is a potential influence on their ratings of acceptability (Rhoades & Kratochwill, 1992). Therefore, another possible explanation for having different acceptability ratings might be that special education teachers are potentially more familiar with the reading intervention in the case scenario than general education teachers. This situation might cause different acceptability ratings on the *RIRP* for teachers.

Finally, the current findings support the conceptual model of Elliott (1988) for treatment acceptability. The model showed that treatment acceptability has complex relationships among various variables/factors and many factors could affect teachers’ acceptability of interventions. The current study corroborates the idea that demographic factors such as years of teaching experience and certification type have an influence on teachers’ acceptability ratings, while extending what is known by indicating that level of
education, regardless of type (i.e., general and special education teacher) may not have an impact on acceptability ratings.

**Limitations**

Despite the potential importance of the findings of this study, there are some limitations to the research that must be taken into account. This section was grouped under three categories: (1) Psychometric, (2) Methodological, and (3) Statistical.

**Psychometric Limitation**

The sample size was small in terms of conducting the EFA for the 15-item RIRP. Researchers generally suggest that there should be at least 10 respondents per item to have a reliable factor structure (Harman, 1976; Schwab, 1978). Also, some researchers recommend that there should be at least a 1:5 item-to-respondent ratio (Fabrigar et al., 1999). Based on these suggestions, the minimum number of participants would be 75 individuals per subgroup. The current study included 59 special education teachers and 78 general education teachers. While sample size for general teachers was adequate, it was small in relation to special education teachers. However, the results from the measures of sampling adequacy indicated that the sample size was sufficient for the 15-item scale used for both groups.

**Methodological Limitations**

I used analogue methodology to investigate the psychometric properties of the RIRP and examine teachers’ acceptability ratings on the RIRP. The approach of using analog methodology did not allow participants to actually implement the reading intervention in a classroom setting. Instead, participants read a fictitious case scenario
and rated the acceptability of the intervention based on the given intervention steps. In other words, teachers’ experience of the interventions did not occur in real settings with real and practiced scenarios. Previous research investigating treatment acceptability has predominantly involved an analogue methodological approach (i.e., Kazdin, 1980; Kelley et al., 1989; Martens et al., 1985; Tarnowski & Simonian, 1992; Von Brock & Elliott, 1987; Witt & Elliott, 1985; Witt & Martens, 1983). However, a small number of studies actually involved implementation of the focal the interventions that were assessed for treatment acceptability (Reimers & Wacker, 1988; Reimers et al., 1991). While there are some advantages to the use of analogue methodologies (i.e., having control on the variables, taking less time, and needing less effort), this approach has limited ecological validity (Miltenberger, 1990). Therefore, the methods used to assess acceptability in this study have limited ecological validity and may not be appropriate to some interventions used in real classroom contexts. More research is needed, clearly, to determine the congruence between analogue and implemented tests for the purpose of rating acceptability of academic interventions.

As indicated above, the sample size was small. The reason of having small sample size might be related to the way of reaching participants. When I was collecting data, I was not able to directly contact with teachers because school superintendents and/or school principals were the people who distributed the survey to their teachers. I could contact very limited number of teachers directly. Even though I sent reminders to school superintendents and/or school principals to distribute the survey, I was not sure if their teachers received the survey.
**Statistical Limitations**

As being a self-report instrument, the *RIRP* has its limitations. Self-report instruments generally exhibit weak internal validity due to the fact that participants may mislead researchers by answering questions in a way that the researchers can more easily accept (Calder, 1998). This case might be a threat to internal validity because participants might have rated the intervention in a way that they believe makes them look more on the side of the researcher. Furthermore, the intervention in the case scenario might be a well-known intervention among teachers because of having simple steps to implement (Huang et al., 2008). This situation also might be a threat to the internal validity because participants might have personal familiarity with the intervention in the case scenario, potentially causing them to have a bias toward (or against) the intervention.

With reference to external validity, this study had a relatively homogeneous sample regarding the ethnicity and gender of the responding teachers. Therefore, the results may not generalize to a more diverse sample of participants. Furthermore, the participants of this study were elementary school teachers (general education and special education) who have experience with struggling readers. As a result, the study findings are only generalizable to similar teacher populations. Additionally, 423 participants participated in the study yet only 137 of them successfully completed the survey. The return rate in this study was not high, as it was around 32%. Thus, the findings may not represent the whole population of teachers.
Implications for Future Research

Many factors contribute to the research-practice gap that affects the implementation of effective practices in schools (Cook et al., 2012). Teachers’ judgments of recommended practices potentially affect their implementation of such practices in classrooms. As such, social validity is particularly important to consider as teachers’ judgments about practices determines their decisions and commitment to implementation (Reimers et al., 1992). Investigating treatment acceptability as a major element of social validity is necessary to understand whether and why recommended practices are acceptable and to determine the factors that have an influence on teachers’ evaluations regarding the use of effective reading practices. Thus, researchers could take one more step to close the research-to-practice gap.

The current study provided preliminary evidence of the psychometric soundness of the RIRP. At present, there are no other empirically validated instruments designed to measure the acceptability of reading interventions. Future research is needed to replicate the factor structure and to examine the reliability and validity of the RIRP. To obtain reliable and consistent data, repeated evaluations will be needed (Chafouleas, Briesch, Riley-Tillman, & McCoach, 2009). Additionally, future research will need to employ a larger sample size to confirm a reliable factor structure for the RIRP.

This study investigated pretreatment acceptability, as rated by participants using the RIRP. The choice of teachers regarding the acceptability of a recommended treatment could happen before, during, and after use of the treatment. The treatment might not be maintained, however, without training and a consideration of the nature of
implementation fidelity and necessary adaptations (State, Harrison, Kern, & Lewis, 2017). Therefore, future investigations should focus on measures of acceptability as rated concurrent (i.e., during implementation) to treatment and after the cessation of treatment and examine whether teachers’ pre-treatment ratings of acceptability, and their concurrent- and post-treatment acceptability ratings are equally predictive for successful implementation of a recommended intervention. Also, participants in the current study completed the RIRP based on one type of reading intervention. Thus, different results could be found if different interventions were used (Chafouleas et al., 2009). Future research should examine the ratings of the participants on the RIRP by providing different interventions.

Future research examining treatment acceptability should also focus on determining other factors that can influence the acceptability ratings of recommended reading interventions. This study investigated some of the teacher factors that could affect teachers’ ratings; however, other researchers have indicated that there are many other factors affecting teachers’ acceptability ratings, such as treatment-related factors (i.e., Kazdin, 1980), contextual factors (i.e., Martens et al., 1985), and student-related factors (i.e., Elliott et al., 1986). All of these factors have been investigated in relation to the acceptability of behavioral interventions, and as such, should also be examined as potential factors influencing teachers’ judgments of recommended academic (e.g., reading) practices. Future research will need to not only include treatment acceptability ratings, but also extend to involve an investigation of the full set of influencing factors when evaluating the effectiveness of reading interventions. Additionally, the issue of
treatment acceptability is just one part of implementing effective reading practices in classrooms. Future research should also expand the questioning of acceptability of treatment procedures to investigating the all components of social validity (e.g. treatment goals and treatment outcomes) in relation to teachers’ judgments of effective reading practices. Researchers should investigate, for example, the acceptability of planned and unplanned outcomes of an effective reading intervention; thus, exploring these other may also contribute to understanding of and ultimately help close the research-to-practice gap.

Finally, rating scales with Likert-type items provide only a limited snapshot of teacher perceptions of the acceptability, as rating scales may overly restrict the answers provided by respondents (Sheatsley, 1983). In this study, participants did not have any opportunity to explain their decisions as they selected from a standard set of options. Future studies may use a variety of data collection methods, such as open-ended questions and interviews. Such data would significantly build on our understanding of why teachers accept or fail to accept effective reading interventions.

**Conclusion**

Treatment acceptability is critical when considering the implementation of effective reading practices in classrooms. Teachers’ judgments regarding recommended practices could affect the use of those practices and outcomes (Cross-Calvert & Johnston, 1990). To date, research has focused mainly on the acceptability of behavioral interventions, and very little research has explored the acceptability of academic interventions. Specifically, there is a need to explore teachers’ perceptions of acceptability for reading interventions. Prior to this study, no empirically validated
instrument had been developed to assess the acceptability of reading interventions, or
explored the factors related to acceptability ratings of teachers related to reading
interventions.

The first objective of this study was to develop a psychometrically sound measure
for determining the acceptability of reading interventions. Thus, I developed and
explored the factor structure of a 15-item reading intervention acceptability measure, the
*RIRP*, by modifying the IRP-15 (Martens et al., 1985) to be suited for evaluating reading
interventions. My second objective was to examine whether teachers’ acceptability of
reading interventions differed by teacher characteristics (i.e., years of experience, level of
education, certification type). For this purpose, I designed, distributed, and scored a
survey containing descriptive information as well as analogue case scenario that provided
information about teachers’ decisions related to the acceptability of reading interventions.

Related to my first objective, the calculation of internal consistency reliabilities
showed that the *RIRP* has sufficient internal consistency and reliability to be of use in
assessments of the acceptability of reading interventions. Findings from an exploratory
factor analysis of the underlying factor structure suggested that the *RIRP* measures one
general acceptability factor for both general education teachers and special education
teachers. Regarding the second objective of the study, my results indicated that teachers’
acceptability ratings of reading interventions differed based on their years of teaching
experience. Findings for both special and general education teachers suggested that
teachers who have fewer years of experience rated the *RIRP* items as being more
acceptable than did those teachers who had more years of experience. However, the
acceptability ratings for both groups of teachers on the *RIRP* did not differ based on their level of education (i.e., bachelor’s degree or master’s degree). Additionally, the findings suggested that the acceptability ratings of the special education teachers on the *RIRP* were higher than those of the general education teachers. Finally, the limitations of the study, the implications of the research findings, and the recommendations for future research were discussed.
APPENDICES
Appendix A

INFORMED CONSENT TO PARTICIPATE IN A RESEARCH STUDY

Study Title: Examination of a Rating Scale to Assess Teachers’ Treatment Acceptability of Reading Interventions for Struggling Readers in Elementary Schools

Principal Investigator: Neslihan Unluol Unal

You are being invited to participate in a research study. This consent form will provide you with information on the research project, what you will need to do, and the associated risks and benefits of the research. Your participation is voluntary. Please read this form carefully. It is important that you ask questions and fully understand the research in order to make an informed decision.

Propose: Effective reading interventions are necessary for students who have problems in reading. However, effective reading interventions are not always used in schools. Research suggests that treatment acceptability (i.e., whether school personnel find the intervention acceptable) is one of the factors that determines whether a particular reading intervention will be adopted and used. The purpose of my study is to take initial steps toward developing a rating scale designed to assess the acceptability of reading interventions. A second major purpose is to investigate whether teachers’ judgments of the acceptability of reading interventions differ by teacher demographics (i.e., teacher type, level of experience, level of education).

Procedures: If you choose to participate in this study, I would like you to complete an anonymous online survey. You will begin this survey by answering basic demographic questions. Then, you will read a brief description of a student with reading difficulties and a reading intervention. Based on this description, you will respond to 15 items using a 6-point rating scale (i.e., 1 = Strongly Disagree, 2 = Disagree, 3 = Slightly Disagree, 4 = Slightly Agree, 5 = Agree, 6= Strongly Agree). The estimated time required to complete the survey is 10 - 15 minutes.

Benefits: This research will not benefit you directly. However, your participation in this study will help us examine and better understand teachers’ judgments of the acceptability of reading interventions.

Risks and Discomforts: There are no anticipated risks involved in participating in this research.
Privacy and Confidentiality: All data gathered from this online survey will be kept confidential. Your name will not be included on the survey to maintain confidentiality. Any identifying information will be kept in a secure location and only the researcher will have access to the data. Research participants will not be identified in any publication or presentation of research result.

Voluntary Participation: Taking part in this research study is entirely up to you. You may choose not to participate or you may discontinue your participation at any time without penalty or loss of benefits to which you are otherwise entitled.

Payment: By completing this survey, you will be eligible to enter into prize drawing for one of the two $50 Amazon gift cards. You will need to provide your e-mail address to contact with you at the end of the survey if you would like to enter into prize drawing.

Contact Information
If you have any questions or concerns about this research, you may contact Neslihan Unluol Unal at 210-544-1473 (e-mail: nunluol@kent.edu) or Andrew Wiley at 330-672-0581 (e-mail: awiley5@kent.edu). If you have any questions about your rights as a research participant or complaints about the research, you may call the IRB at 330-672-2704.

If you understand the statements above and freely consent to participate in this study, click on “I Agree” button to start the survey.

I Agree  I Do Not Agree
APPENDIX B

READING INTERVENTION RATING PROFILE (RIRP)
Appendix B

READING INTERVENTION RATING PROFILE (RIRP)

Based on the description you just read, please select the number that best describes your agreement or disagreement with each statement.

1 = Strongly Disagree  2 = Disagree  3 = Slightly Disagree  4 = Slightly Agree  5 = Agree  6 = Strongly Agree

<table>
<thead>
<tr>
<th>Statement</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. This would be an acceptable intervention for the child’s reading problem.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>2. Most teachers would find this intervention appropriate for reading problems in addition to the one described.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>3. The intervention would increase the child’s reading skills and achievement.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>4. I would suggest the use of this intervention to other teachers.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>5. The child’s reading problem is severe enough to warrant use of this intervention.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>6. Most teachers would find this intervention suitable for the reading problem described.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7. I would be willing to use this in the classroom setting.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>8. The intervention would not result in negative side-effects for the child.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>9. The intervention would be an appropriate intervention for children at varying levels of reading skills.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>10. The intervention is consistent with those I have used in classroom settings.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>11. The intervention was a fair way to handle the child’s reading problem.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>12. The intervention is reasonable for the reading problem described.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>13. I like the procedures used in the intervention.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>14. This intervention was a good way to handle the child’s reading problem.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>15. Overall, the intervention would be beneficial for the child.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
APPENDIX C

DEMOGRAPHIC QUESTIONS
Appendix C

DEMOGRAPHIC QUESTIONS

Please read the following questions and choose the appropriate answer that best describes you.

1. What is your age? ___ (Please type your age in years here and round to the nearest whole number. For example, if your age is 33 years and 8 months, you would type “34” here.)

2. What is your gender?
   __ Male
   __ Female

3. Which race/ethnicity best describes you?
   __ White / Caucasian
   __ Hispanic / Latino
   __ Black / African American
   __ Native American / American Indian
   __ Asian / Pacific Islander
   __ Native Hawaiian / Other Pacific Islander
   __ Multiple Races
   __ Other (Please specify)

   [ ]
4. Please indicate the type and/or highest level of training you have received in your area of teacher licensure. If you have training other than teacher licensure, please specify your type of training under the “Other” section.

- Bachelors Degree
- Masters Degree
- Education Specialist (Ed.S.)
- Doctor of Education (Ed.D.)
- Doctor of Philosophy (Ph.D.)
- Other (Please Specify)

5. How many years have you worked as a teacher? Please specify.

6. Have you worked with struggling readers?

- Yes
- No

7. Please select your teaching certification.

- Early Childhood
- Middle Childhood
- Adolescent Education
- General Education
- Special Education (Mild to Moderate)
__ Special Education (Moderate to Intensive)

__ Other (please describe)

8. What grade or grades do you currently teach?

9. What reading program(s) and/or instructional approach/approaches do you use in your class or grade? (e.g. Balanced Literacy, Accelerated Reading, Reading Recovery, Success for All)

10. Has your school where you are currently working implemented Response to Intervention (RtI)?

__ Yes

__ No

__ Other (Please Specify)
Appendix D

INTERVENTION CASE STUDY

Please carefully read the following description of a student and a reading intervention. Then, answer the questions in the following questionnaire based on this description.

Danny attends a regular class in a public school. Danny has difficulty reading fluently. He reads slowly and haltingly, and without appropriate expression. Danny’s difficulties with fluency interfere with his ability to understand what he is reading. Informal assessments indicate that Danny is below grade level in reading. Therefore, Danny’s teacher decides to implement a reading intervention to address Danny’s reading difficulties. When implementing the reading intervention, Danny’s teacher removes all distractions (e.g. pencils, toys, or objects) from the setting and makes sure that she has all the necessary materials ready.

The intervention steps are as follows:

INTERVENTION STEPS

1. Danny’s teacher reviews rules and expectations related to staying on task and following teacher directions.

2. To increase comprehension, Danny’s teacher briefly previews the passages that Danny will read as part of the reading intervention.

2. Danny’s teacher positions the selected instructional-level reading passage for the session so that both Danny and his teacher easily follow the text.

3. Danny’s teacher says, “Read aloud the given passage as quickly as you can while being careful to understand what you read.” The teacher secretly keeps time using the stopwatch and marks words read incorrectly on a duplicate copy of the passage. The purpose is to measure fluency (i.e., the number of words read correctly per minute)

4. When Danny misreads a word or pauses longer than 3 seconds, Danny’s teacher reads the word aloud, asks Danny to repeat the word, and has him to continue reading.

5. When Danny is finished reading, Danny’s teacher records the elapsed time and the number of correct words (to allow calculation of words read correctly per minute).
6. Danny’s teacher has him to reread the same passage two to four times until he meets 85% word accuracy. Danny’s teacher records the number of correct words per minute and number of incorrect words and provides corrective feedback for each passage reading.

7. Danny’s teacher praises him for his efforts and successes. Additional passages are read as time allows.
Appendix E

EXPERT EVALUATION SHEET

Dear Expert Reviewer,

Thank you for your willingness to assist with the development of the Reading Intervention Rating Profile (RIRP).

The RIRP measures teachers’ perceptions of the treatment acceptability of different reading interventions. A copy of the RIRP and a review of research on treatment acceptability are included for your convenience. Please provide your feedback, comments, and suggestions on the items of the instrument. See the following questions that help you to review the instrument:

- Do the items clearly relate to the construct being measured (treatment acceptability of reading interventions)?
- Are the terms and language used in the rating scale current, clear, and understandable?
- How the items need to be rephrased if necessary?
- Do you have any other suggestions for improving this rating scale?

At your earliest convenience, please email your review to Neslihan Unluol Unal at nunluol@kent.edu

If you have questions, do not hesitate to contact me and/or my advisor (Dr. Andrew Wiley, awiley5@kent.edu). Again, thank you so much for your assistance!

Neslihan Unluol Unal, Doctoral Candidate
Special Education, Kent State University
References


Lubke, G., & Muthén, B. (2004). Factor-analyzing Likert scale data under the assumption of multivariate normality complicates a meaningful comparison of observed groups or latent classes. Structural Equation Modeling, 11, 514-534.


Mastropieri, M. A., & Scruggs, T. E. (2005). Feasibility and consequences of response to intervention examination of the issues and scientific evidence as a model for the


challenges at the high school level. *Journal of Positive Behavior Interventions, 19*(1), 26-36.


http://jcmc.indiana.edu/vol10/issue3/wright.html


