5-2010

Exploring the Relationship Between Factors of Implementation, Treatment Integrity and Reading Fluency

Kira Henninger
University of Massachusetts Amherst

EXPLORING THE RELATIONSHIP BETWEEN FACTORS OF IMPLEMENTATION, TREATMENT INTEGRITY AND READING FLUENCY

A Dissertation Presented

by

KIRA LIESE HENNINGER

Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

MAY 2010

School Psychology

EXPLORING THE RELATIONSHIP BETWEEN FACTORS OF IMPLEMENTATION, TREATMENT INTEGRITY AND READING FLUENCY

A Dissertation Presented

by

KIRA LIESE HENNINGER

Approved as to style and content by:

_____
John M. Hintze, Chair

_____
Amanda M. Marcotte, Member

_____
Maureen Perry-Jenkins, Member

_____
Christine B. McCormick, Dean
School of Education

DEDICATION

To my parents, who have shown me nothing but love and support from the very beginning and have always been constant reminders of what is most important in life.

ACKNOWLEDGEMENTS

There are many individuals I would like to thank for supporting me in this project and throughout my educational career. First, I would like to thank my committee members, especially my advisor and chair, John Hintze, for their feedback, support and guidance. I would like to thank Maureen Perry-Jenkins for always being a supportive, caring supervisor and for encouraging me to apply for Graduate School. Thank you to my colleagues, especially Melena Fontaine, Lisa Fisher and Hadley Moore for all the great moments we shared. The laughter and support you provided helped me get to where I am today. I am very grateful for your friendship.

Thank you to everyone who contributed to this study. I want to thank the principal, teachers, instructional aides, and students who participated in the study. I also want to thank Robert Weires, the director of Psychological Services, for sponsoring my dissertation and allowing me to conduct research within the school system. I want to thank Laura Kavlie, Jessica Amargo, Dawn Witchell, Juan Simon, and Akua Darkwah for helping with data collection. A special thank you goes to my intern supervisors, Bridget Theakston and Syndi Olinger, who helped make my dissertation a priority. I am so grateful for your support and guidance along the way, but most of all for your friendship.

I especially want to thank my family. I am truly lucky to have such a supportive, loving network of people who have never given up on me. Finally, I would like to thank John Cartledge, who has always provided me with love and support. You provided me with the motivation to keep going, especially toward the end when I needed it most. Most importantly, you reminded me each day of what was truly important and I am forever grateful for everything you have given me.

ABSTRACT

EXPLORING THE RELATIONSHIP BETWEEN FACTORS OF

IMPLEMENTATION, TREATMENT INTEGRITY AND READING FLUENCY

MAY 2010

KIRA HENNINGER, B.A., UNIVERSITY OF MASSACHUSETTS AMHERST

M.Ed., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor John M. Hintze

Treatment integrity has always had a presence in research, but now more than ever must become a priority owing to the changes in Special Education Law. The present study intends to explore the relationship between factors of implementation, treatment integrity of intervention implementation, and reading fluency. Participants included students in grades 2 through 5 and their teachers enrolled in an urban elementary school in the southwest area of the United States. Participants were chosen for possible inclusion on the basis of their fall performance relative to oral reading fluency on a universal screening measure used as part of the district's Response to Intervention (RTI) plan. Classroom teachers were observed implementing reading interventions and asked to respond to surveys aimed at summarizing their opinions regarding factors related to choice of intervention and implementation. Path analysis was conducted to explore the relationship between two factors of implementation (intervention complexity and

acceptability), treatment integrity (adherence to intervention protocol) and student outcomes (oral reading fluency scores). It was hypothesized that low scores for intervention complexity would be inversely related to levels of treatment integrity, which would subsequently be positively related to reading fluency. Moreover, it was hypothesized that intervention acceptability and treatment integrity would be positively related, which would subsequently be positively related to reading fluency. Lastly, it was hypothesized that there would be an inverse relationship between intervention complexity and reading fluency, and a positive relationship between intervention acceptability and reading fluency. Results indicated an inverse relationship between intervention complexity and treatment integrity, suggesting that when complexity was low, treatment integrity was high. A positive relationship was found between intervention acceptability and treatment integrity, suggesting that when acceptability was high, treatment integrity was high. Furthermore, when treatment integrity was high, reading fluency scores were found to be high. An inverse relationship was found between complexity and reading fluency, suggesting that when complexity was low, reading fluency scores were high. Lastly, a positive relationship was found between acceptability and reading fluency, suggesting that when acceptability was high, reading fluency scores were high.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS……v

ABSTRACT……vi

LIST OF TABLES……...xi

LIST OF FIGURES……xii

CHAPTER

1. STATEMENT OF THE PROBLEM……....1

Treatment Integrity……...1

Operational Definitions…….....3

Special Education Law…….....4

Treatment Integrity in Practice……..…...6

Response to Intervention vs. Traditional Model….……..……7

Procedural Integrity…….……10

Summary……..11

2. LITERATURE REVIEW……..……..13

Technical Issues of Treatment Integrity……..14

Specificity of Definitions……....14
Treatment Dosage……...16
Treatment Exposure and Adherence……...17
Treatment Components……...18

Factors of Treatment Integrity…….18

Complexity……..19
Resources: Time, Materials, and Agents…….21
Perceived and Actual Effectiveness……23
Treatment Acceptability……..24

Assessment of Treatment Integrity…….30

Direct Measures…….30
Indirect Measures……...32

Summary……34

Purpose of the Current Study…….34
Research Questions……39

3. METHODOLOGY……40

Setting and Participants…….40

Setting……40
Participants…….41

Procedure……42

Response to Intervention – District Model…….43
Response to Intervention – Participating School……45

Independent Variables……47

Tier II Interventions….....47
Factors of Implementation……..51

Complexity Factors…….51
Acceptability of Intervention Factors…….52

Measures……...……..52

Acceptability of Intervention Measure……52
Intervention Integrity Checklist–Observations..……..53
Factors of Implementation Checklist–Self Report….....54
Factors of Implementation Checklist–Observations.....55

Dependent Variable……..55

Importance of Reading…….55
Dynamic Indicators of Basic Early Literacy Skills……..57

4. RESULTS……....59

Purpose of the Study/Hypothesis Testing……59
Descriptive Statistics……...60

Inter-observer Agreement……..…....64
Interpretation: Path Analysis Model…….……64

5. DISCUSSION…….....66

Introduction to the Problem……66
Overview of the Study……67
Interpretation of Analysis……...67

Question 1: Direct Effects……..67
Question 2: Indirect Effects……73

Limitations……..74
Implications for School Psychologists……....77
Directions for Future Research……...79
Summary……80

APPENDICES

A. DISTRICT INTERVENTION PROTOCOL…….....……...…...82
B. DISTRICT INTERVENTION PLAN…....……....…….83
C. DISTRICT INTERVENTION LOG…..……..84
D. DISTRICT INTERVENTION GRAPH……..85
E. ACCEPTABILITY OF INTERVENTION MEASURE (AIM)–PRE-INTERVENTION SURVEY…….……..86
F. INTERVENTION INTEGRITY CHECKLIST……...……...87
G. ACCEPTABILITY OF INTERVENTION MEASURE–POST-INTERVENTION SURVEY……...……88
H. FACTORS OF IMPLEMENTATION CHECKLIST–SELF REPORT MEASURE…….……89
I. FACTORS OF IMPLEMENTATION CHECKLIST–OBSERVATIONS…92

REFERENCES……94

LIST OF TABLES

Table Page

1. Univariate Statistics……...……61
2. Complexity Subscales: Means and SD's…….……..62
3. Ranges: Scales and Subscales……62
4. Bivariate Correlations……63

LIST OF FIGURES

Figure Page

1. Conceptualization of Witt and Elliott's (1985) proposed model of treatment acceptability. ……...……26
2. Path Analysis Model…….65

CHAPTER 1

STATEMENT OF THE PROBLEM

Treatment Integrity

The purpose of theory development is to explain natural phenomena, seek and explain links among a range of behaviors with a focus on understanding, prediction, and control (Heppner, Kivlighan, & Wampold, 1999). Owing to a lack of treatment integrity data, however, in conducting research many researchers cannot say whether their experiments have proved or disproved their hypotheses. Treatment integrity is the degree to which an independent variable or intervention is implemented as intended (Gresham, 1989). Treatment integrity examines whether the intervention agent implements an intervention as intended and measures the difference between what is expected and what is actually executed.

Unfortunately, little attention has been paid in research to the measurement of treatment integrity. Too often, treatment integrity is assumed rather than measured. Several researchers have reviewed the literature to assess exactly how frequently integrity was measured and found that only 15% to 20% of articles measured and reported integrity data and only 16% to 35% provided an operational definition for the independent variable (Gresham, Gansle, Noell, Cohen & Rosenblum, 1993; Peterson, Homer & Wonderlich, 1982). Gresham et al. (1993) reviewed 181 experimental studies from seven journals known for behaviorally based interventions between 1980 and 1990, looking for studies that included treatment integrity data, operational definitions of treatments, and reporting of effect size. The purpose was to update and expand upon the findings from the Peterson et al. (1982) review of reported treatment integrity data in 539

studies published by the *Journal of Applied Behavior Analysis* between 1968 and 1980. Integrity of independent variables in each study were measured and placed into one of three categories: 1) yes, involving inter-observer agreement on application of the independent variable or calibration checks by the researcher; 2) no, but the risk of inaccurate application was low; and 3) no integrity was assessed, but it was not necessary. Peterson et al. (1982) found that only 20% of the 539 studies reported treatment integrity data. Operational definitions of the independent variables were also measured and placed into one of three categories: 1) yes, 2) no, but unnecessary and 3) no, but necessary. Only 16% of the studies reported operational definitions of the independent variable.

Gresham et al. (1993) expanded upon the existing literature by including stricter criteria for determining whether evidence of treatment integrity or operational definitions was presented in each study. The six criteria for inclusion were: 1) an experimental study that assessed the effect of treatment on behavior; 2) participants under the age of 19 years; 3) publication date between 1980 and 1990; 4) enough information provided to compute effect size; 5) exclusion of case studies and group designs without control groups that did not allow for causal inference; and 6) experimental studies must have taken place in a school. Graduate students reviewed the 181 articles matched to the above criteria and coded the following variables: subjects, treatment, dependence, design, and effect size. Age level was not significantly related to any other variable. Results indicated that 14.4% (26 studies) measured and reported integrity data and 34% (65 studies) provided operational definitions of treatment. Treatment integrity data were considered to be reported if a percentage of treatment integrity was provided.

Operational definitions were considered to be provided if enough information was given (i.e. specific verbal, physical, temporal, and/or spatial parameters of the independent variable) to facilitate replication (Gresham et al., 1993). Further discussion of operational definitions is addressed in the following section.

Approximately 75% to 80% of studies not measuring treatment integrity data indicate that the functional relationship determined between dependent and independent variables had clearly been compromised in the majority of the literature (Gresham et al., 1993). When treatment integrity data are not provided, the researcher lacks support in identifying the independent variable as the cause of the changed behavior. Without treatment integrity data, one cannot determine whether the fact that an intervention was unsuccessful in changing behavior was due to an ineffective intervention or to a low level of implementation integrity. To remedy this, it must be demonstrated that changes in a dependent variable (behavior) are related to systematic changes in an independent variable (i.e. the intervention) through the measurement and reporting of treatment integrity (Gresham, 1989).

Operational Definitions

Popper (1994) stressed the importance of hypothesizing, criticizing, and then revising when conducting research. Popper's contribution to research was the notion of falsifiability, which states that the primary characteristic of a theory is that it can be replicated and tested again. Researchers must not only seek to support a theory but, more importantly, seek to refute a theory. Gresham (1996) notes that research cannot be replicated if the researcher does not describe in detail what was done, how it was done, and for how long it was done. In not providing adequately defined operational

definitions of variables, other researchers are not able to replicate the study and therefore are not able to provide evidence to support or refute the theory. Gresham et al. (1993), as stated above, found that only 34% of studies reported operational definitions of the variables in a review of treatment integrity, indicating that the majority of the research being published cannot be tested to support or refute the findings. Operational definitions were referred to by Kerlinger (1986) as "a sort of manual of instructions" that provide the necessary procedures allowing for replication of the study. By not providing operationally defined variables, therefore, researchers make it difficult for others to replicate the research and thereby do not provide a testable theory. Operationally defining behavioral components of a study allows for clear measurement of the desired behavior and strengthens the study's construct validity. Construct validity refers to how well the variables represent the constructs that they are intended to measure (Heppner et al., 1999). Variables that are operationally defined in specific, measurable terms can be monitored easily to facilitate the ruling out of other potential confounding variables in the study that may compromise the results (Gresham, 1996). Operational definitions of variables facilitate the measurement of treatment integrity, while collection of treatment integrity data allows the researcher not only to monitor the implementation of the variable but also to monitor other possible confounding variables that may have an affect on the results.

Special Education Law

Measurement of treatment integrity is necessary not only for facilitating good research practice but for complying with special education law. Measuring the treatment integrity of intervention implementation is beginning to have an even greater importance

in conducting research, owing to the increased focus of Response to Intervention (RTI) in schools. Public Law 108-446, more commonly referred to as the reauthorization of the Individuals with Disabilities Education Improvement Act (IDEA 2004), which took effect on July 1, 2005 (U.S. Department of Education, 2004), was the first legislation to introduce the option for an alternative method of determining eligibility for specific learning disabilities. The current law states the following with respect to specific learning disabilities [IDEA; 614,b,6,A,B].

> (A) IN GENERAL – Notwithstanding section 607(b), when determining whether a child has a specific learning disability as defined in section 602, a local educational agency shall not be required to take into consideration whether a child has a severe discrepancy between achievement and intellectual ability in oral expression, listening comprehension, written expression, basic reading skill, reading comprehension, mathematical calculation, or mathematical reasoning.
>
> (B) ADDITIONAL AUTHORITY – In determining whether a child has a specific learning disability, a local educational agency may use a process that determines if the child responds to scientific, research-based intervention as a part of the evaluation procedures described in paragraphs (2) and (3).

Special education law has evolved in allowing other methods for the identification of students with a specific learning disability (SLD), owing to the dramatic increase in identification over the past several decades. Fifty percent of all students with disabilities have a SLD (U.S. Department of Education, 1999). In accordance with law, RTI has become an accepted alternative model of identifying students with a SLD. Although

special education law dictates its use for identification, RTI is a general education initiative that benefits all students. Although an RTI framework may be applied to all academic subjects, as well as behavior, there has been a primary focus on reading as it is an essential skill for all academic subjects. Research has shown that approximately 75% of students with reading problems in third grade continue to have difficulty with reading in ninth grade (Francis, Shaywitz, Stuebing, Shaywitz & Fletcher, 1996), which supports the primary focus on reading.

Treatment Integrity in Practice

RTI is not any one particular intervention but a preventive model based on several core principles used to provide high-quality instruction and interventions with a sound research base that are matched to student need. It consists of three main components that provide its framework: 1) a multi-tiered system to provide service delivery; 2) a psychometrically sound assessment system to measure students' progress; and 3) a systematic problem-solving process (National Association of State Directors of Special Education, NASDE, 2005). RTI is built upon the premise that instruction and interventions are implemented with integrity. The success of RTI as an acceptable model for identifying a SLD depends upon treatment integrity, as a student's response to an intervention implemented with treatment integrity is the primary factor determining the level of intervention required (National Association of State Directors of Special Education, NASDE, 2008). Gresham (2009) addresses the role that treatment integrity plays within a RTI framework; "One cannot legitimately claim to practice response to intervention in the absence of determining the extent to which the 'I' (intervention) was

implemented" (p. 538). Essentially, there is no RTI without treatment integrity of intervention implementation.

Noell and Gansle (2006) dictate the need for assessment and documentation of treatment integrity in order to avoid violation of due process rights provided to students in an RTI model. If a scientifically based intervention is not implemented with integrity, one can no longer infer that the same results produced during the research phase of its design can be expected. For example, the Read 180 reading intervention is designed to be conducted in a whole-class format. There are two elements: formal teacher-led instruction time and a computer-based program that must be implemented five days per week for a 90-minute time block. If a teacher decides to use Read 180 for a student in her class using only the computer component for 20 minutes per day, the results yielded in research development cannot be expected and therefore the teacher cannot claim that an evidence-based intervention was provided. Adherence to an intervention protocol is essential, therefore, as outlined in current legislation.

Response to Intervention (RTI) vs. Traditional Model

RTI relies on a dual-discrepancy model that requires educators to provide evidence that the student both performs below the level shown by classroom peers and demonstrates a learning rate over time substantially slower than that of classroom peers (Fuchs, Fuchs, & Speece, 2002) rather than the traditional model (i.e. intelligent quotient/achievement discrepancy model), which relies solely on student performance in comparison to peers for the identification of a SLD. One of the most well-known limitations of the discrepancy model has to do with the over-representation of minority students (Fuchs, Mock, Morgan, & Young, 2003), which further supports the case for

using more than one dimension for determining eligibility. One datum point in time is not sufficient for identifying a student with a SLD and does not indicate a need for special education services. Fuchs et al. (2002) provide an excellent example of how pediatric medicine uses the same principles of the dual-discrepancy model and RTI when determining need for intervention. When a child's height is measured, at a routine office visit, to be below the third percentile, the doctor does not initially assume the possibility of an underlying pathology but measures and monitors the child's growth over time and compares the child's growth rate to the growth rate of typically developing peers (Fuchs et al., 2002). If the child's rate of growth is commensurate with that of peers, despite being at the third percentile in height, the child is clearly able to successfully derive benefits from the environment and is not a candidate for intervention. In education, the same principles apply. A student performing at a lower level than same-age peers does not necessarily indicate a learning disability if the student is able to learn at a similar rate as peers. Demonstration of progress over time indicates that the student is benefiting from the educational environment and that special education intervention is therefore not necessary.

The traditional discrepancy model is reactive and has been referred to as a "wait to fail" model (Brown-Chidsey & Steege, 2005), delaying needed support. Children demonstrating a discrepancy receive additional instruction but only after weeks or months of struggling. More importantly, children that do not meet the discrepancy criterion but demonstrate both low cognitive skills and poor academic ability are still left struggling without the support necessary for them to succeed educationally. Limitations of the traditional model that have been identified have increased the focus on the

development of an alternative model. An RTI model is a data-based, prevention-focused system that matches instruction to student need, providing all students with evidence-based instruction and intervention (Wedl & Schroeder, 2005). The multi-tiered system of service delivery, combined with psychometrically sound assessment tools, provides a framework for identifying student deficits early and necessary interventions specific to need in order to prevent future difficulties.

The traditional model evaluates the discrepancy between a child's scores on a cognitive measure and an achievement assessment. For a child to be eligible according to the traditional model, the score on the achievement test needs to be approximately 12 points lower than the score received on the cognitive test. It is "approximately" 12 points lower owing to the variation in law among states, which is one of many limitations of the traditional model – it lacks consistency. Furthermore, the traditional model is built on the assumption that cognitive and achievement test scores are equal. The problem here is that, for this assumption to be true, the two tests would need to be completely correlated. However, there is only a correlation, of about 0.6 (Sattler & Dumont, 2004). In addition, when these tests are used together, the individual reliability of each test decreases from 0.9 to a compounded reliability of about 0.75 (Sattler & Dumont, 2004). Martson (1989) criticizes the psychometric properties of cognitive and achievement assessments and the lack of reliability and validity, particularly construct validity data to support use in identifying children with a SLD. Also, cognitive and academic assessments are not linked to intervention and are not able to be used formatively to aid in decision-making (Martson, 1989). The traditional model uses indirect measures of tracking to understand student performance (Martson, 1989), providing little or no information for intervention

or instructional design. Additionally, fluency (Martson, 1989), which the National Reading Panel found to be one of the five building blocks of reading, is not measured.

Included in RTI practices is a psychometrically sound set of assessment tools that are short, quick, inexpensive, and easily administered. These are referred to as Curriculum-Based Measurement (CBM). The scores they produce are used as a "performance indicator" representing an individual's global competence in a domain of interest (i.e. reading fluency, reading comprehension, mathematics, written expression, or spelling) (Deno, 1989). According to several studies (Buck & Torgeson, 2003; National Center for Student Progress Monitoring, NCSPM, 2008), CBM continues to show strong psychometric properties in the areas of reliability and validity. In addition, CBM is a direct measure of student skills that provides alternative forms to be used for formative assessment (National Center for Student Progress Monitoring, 2008). Each CBM probe is easily administered and scored following standardized methods and provides useful information in terms of a student's skill in that particular area. CBM tools used in a RTI framework target specific skills that provide essential information about a child's deficits, leading directly to the type of intervention necessary. When these assessment tools are used formatively, student progress can be measured to indicate whether the intervention is working or whether adjustments need to be made to instruction.

Procedural Integrity

Glover and DiPerna (2007) note that "procedural integrity" incorporates both adherence to an established protocol for the provision of services (e.g. number of intervention tiers, decision-making criteria) as well as integrity of intervention provision within each tier of services. It represents another key component for RTI models of

service delivery. Procedural integrity measures how well the procedures or steps of a framework are followed and is equally as important as treatment integrity. When procedural integrity is adhered to, the implementation of the provision of services and interventions within each provision are followed as originally intended. There are multiple facets of the RTI model in which a breakdown or lack of procedural integrity may occur. Within the multi-tiered system of service delivery, procedural integrity may be threatened in the following procedures: implementation of Tier I core instruction, implementation of Tier II interventions, implementation of Tier III interventions, data collection, communication among team members, and adherence to established decision-making procedures. School psychologists and other educators must make efforts to ensure that procedural integrity is a priority in the implementation of outlined procedures as well as the application of interventions if they hope to foster the development and implementation procedures of RTI as outlined in educational law. Glover and DiPerna (2007) identify maintenance of procedural integrity as one of the primary components of RTI. Future research of procedural integrity should focus on the areas of acceptability, training, and support.

Summary

Treatment integrity research has largely been ignored over the past several decades. Recent attention to the overall lack of treatment integrity data reported in research and current legislative support for the use of the RTI process indicates a greater focus on treatment integrity in both research and practice. Legislative support and contradictory research findings for the success of the traditional model for a SLD determination have pressed RTI to the forefront as a way for schools to approach students

who are struggling to learn. Evidence that the traditional model is reactive, lacks use of psychometrically sound and relevant assessments, and provides no link between assessment and instruction has demonstrated a need for an alternative model. In contrast, RTI can be preventive, uses psychometrically sound and relevant assessments, and uses assessment to drive instruction, making it a more desirable option not only for the determination of SLD eligibility but also for a general education framework focused on serving all students.

Gansle and Noell (2007), in a synthesis of treatment integrity research, concur that treatment integrity, or what they refer to as "treatment plan implementation" is fundamental to the success of RTI and therefore additional research on the topic is a major priority. To ensure the successful implementation of RTI and all of its key components, including implementation of scientifically evidenced interventions, measuring treatment integrity is an essential practice deserving more attention.

CHAPTER 2

LITERATURE REVIEW

The purpose of this chapter is to review and critically summarize literature in the following areas of treatment integrity: technical issues, factors of implementation, and assessment. Many questions have been raised in the literature on how treatment integrity should be measured. Questions surrounding technical issues in treatment integrity research include specificity of definitions, treatment dosage, treatment exposure and adherence, and treatment components. Several factors of implementation have been identified as relating to treatment integrity. Gresham (1989) identified six factors, including complexity, time, materials and resources, number of treatment agents, and perceived and actual effectiveness. Acceptability by the intervention agent is another factor noted to influence treatment integrity by several researchers (e.g. Eckert & Hintze, 2000; Elliott, Witt, & Kratochwill, 1991). These factors must be explored more deeply to provide answers to specific questions surrounding their importance in treatment integrity. For example, how much treatment is necessary to elicit the intended effect? What is the level of complexity at which an intervention agent is less likely to implement an intervention with treatment integrity? What defines the complexity of an intervention? Finally, how should treatment integrity be measured? Should it be assessed directly through observations or indirectly through rating scales, interviews, and self-reporting? There are both advantages and disadvantages to direct and indirect types of assessment, but what is the most accurate way to measure how well treatments are being implemented in the classroom? These questions and issues are addressed in the next section.

Technical Issues of Treatment Integrity

Specificity of Definitions

Treatment components must be defined so that they can be measured accurately. One technical issue related to the measurement of treatment integrity deals with how specifically components should be defined. Much attention in research has been devoted to defining measured variables (i.e. dependent variables) but much less attention has been paid to defining independent variables, as noted previously by Gresham et al. (1993) and Peterson et al. (1982), which also provide support to the importance of providing operational definitions. Should the treatment components be defined globally, intermediately, or molecularly (Gresham, 1996)? If the treatment component is defined more globally, it is possible that it is not fully being captured and will reduce the validity and reliability of the observational system (Epps, 1985; Skinner, Rhymer, & McDaniel, 2000). If the treatment component is defined too specifically, it is possible that the teacher will not implement the program as designed or at all, decreasing treatment integrity because it is too complex or time-intensive. An example of how dependent and independent variables are defined globally, intermediately, and molecularly follows.

There are several ways to define "reading" as a behavior (dependent variable) when discussing the expectations of the student. "Johnny will read one passage each week with fluency and accuracy" is one way to define the behavior of "reading" globally. In this example, when the behavior is defined globally, it makes it very difficult for the teacher to measure the behavior. The teacher might ask questions such as, "what passage do I use to measure Johnny's reading behavior?", "how do I measure accuracy and fluency?", and "how do I know whether it is accurate or fluent enough?" When the

definition of the behavior is too broad, the behavior is unlikely to be captured. It is more likely that the intervention will be implemented with integrity when the expectations for the teacher are clear, simple, and easily measured and the behavior is defined intermediately. A better example for defining reading behavior is: "Johnny will read one randomly selected Aimsweb Oral Reading Fluency passage once a week. It is expected that he will read 25 words correctly per minute and make no more than 3 errors". In this case, reading behavior is clearly defined and the teacher's expectations for measurement tools are simple and direct. Defining the behavior too specifically or at a molecular level might be defined as above, with the additional requirements that the behavior is measured daily and a detailed list of mistakes is made on each probe.

Typically, operational definitions of independent variables are rare in the literature. It is just as important to adequately define the intervention for the measurement of treatment integrity, as well as for replication purposes. An example of a globally defined protocol for a reading intervention (independent variable) might be, "practice the following reading components during intervention time: phonics and phonemic awareness". Defined at an intermediate level, the protocol might include: 1) introduce the intervention topic of the day; 2) review the sound of the day (e.g. a short "a" sound); 3) read the list of short "a" words aloud together; and 4) have the student read the story aloud. If the behavior was defined too specifically, there would be several sub-steps within each step mentioned above. For example, step #2 would be further divided into: a) have the students say the letter "a"; b) have the students say the sound of the letter "a"; c) have the students write the letter "a"; and d) repeat steps a–c 10 times. The more

narrowly the definition of the independent variable is defined, the lower the treatment integrity is expected to be.

Specificity of definitions is an area needing further research to determine the level of specificity needed when defining the behavior to be measured (i.e. dependent or independent variable). Providing a better understanding in this area will have implications for both research and practice.

Treatment Dosage

Treatment dosage refers to the amount of treatment. Gansle and Noell (2007), in their review of treatment integrity research, found that questions still remain about how much treatment is necessary for an intervention to be successful. Gresham (1996) questioned how much an intervention agent can deviate from an intervention protocol and still get the intended effect of the behavior. There is general agreement that as treatment integrity decreases, treatment effectiveness decreases as well. Nevertheless, research in this area is needed to determine the specific relationship between level of deviation from an intervention protocol and outcomes. Deviation from the protocol may elicit several possible results: positive, negative, or neutral (Gresham, 1996) and there is a need to understand the level of deviation and its association to each of those results.

The importance of treatment dosage is demonstrated through an example related to medicine. When a person has a headache, the recommended dosage of medicine (i.e. pain reliever) is two tablets. If the person takes one tablet instead of two, he/she may experience one of the following results: 1) the headache goes away (positive) or 2) the headache remains (negative). The recommended dosage is two tablets because that is what research has shown as the necessary to get the desired effect. Awareness of

treatment dosage provides intervention agents with necessary information so that just enough dosage is provided to elicit the desired effect. Further research should focus on determining dosage by providing several different levels of the intervention and monitoring the point at which the intervention reached its desired effect.

Treatment Exposure and Adherence

Treatment exposure is the amount of time the student receives instruction on the intervention. Treatment adherence is the extent to which implementation of the independent variable matches the prescribed intervention (Gresham, 1996) or, in other words, how closely the intervention protocol is followed. Typically, exposure is measured in hours and/or days of intervention implementation (i.e. 3 days per week, 1 hour each day) and adherence is measured as a percentage of treatment steps completed (i.e. 7 of 10 steps = 70%) (Schulte, Easton, & Parker, 2009). It is unknown how closely an intervention plan needs to be followed in order to still gain successful results (Gresham, 1989). Owing to measurement issues, researchers have demonstrated difficulty in measuring both treatment exposure and treatment adherence. A self-report measure, for which the intervention agent indicates time of implementation and adherence to intervention steps, may be inaccurate and unreliable owing to bias. It is not feasible to conduct observations for each implementation session, and it is unknown how many observations are necessary to give an accurate depiction of treatment exposure and adherence. The reader is referred to the section on assessment of treatment integrity for a more in-depth discussion of the advantages and disadvantages of each type of assessment data.

Treatment Components

The possible weighting of individual treatment components has yet to be determined (Gresham, 1996). In measuring treatment integrity, it is not known whether there are certain components that are more important to intervention success. It is likely that certain steps in an intervention are more crucial to its success and therefore deserve higher priority when evaluating treatment integrity. It is necessary to identify these steps so that unnecessary components can be removed and time efficiency increased. There has been no research based on the determination of treatment components or on those components that are linked to greater success. Gresham (1996) suggests that research should focus on determining which treatment components demonstrate a functional relationship to the target behavior and give more weight to those components. Further research should explore this functional relationship by manipulating each component within the intervention. Each issue raised above is important in identifying successful ways for measuring treatment integrity and should be considered when conducting future research.

Factors of Treatment Integrity

In evaluating treatment integrity, several factors relating to the level of integrity with respect to intervention implementation described in the literature deserve more attention. Gresham (1989) introduced six factors to be considered in conducting integrity research: 1) complexity of treatments; 2) time required for implementing treatments; 3) materials/resources required for treatments; 4) number of treatment agents required; 5) perceived and actual effectiveness of treatments; and 6) motivation of treatment agents.

Also, Witt and Elliott (1985) introduced acceptability of treatment by the intervention agent as another factor related to treatment integrity.

Complexity

The complexity of the treatment, which has been directly related to the degree of treatment integrity (Yeaton & Sechrest, 1981), deals with the specific components within an intervention. It is generally expected that the more complex the treatment, the less likely the intervention will be implemented with integrity. When looking at the effects of a highly complex Direct Instruction model requiring the implementation of several interrelated treatment components, such as rapid pacing of instruction and frequent questioning, Becker and Carnine (1980) found that treatment integrity was low. The Direct Instruction model emphasizes the use of small-group, direct, face-to-face contact with the teacher and provides systematic and sequenced instruction. Becker and Carnine (1980) summarized major findings from Project Follow Through, which was an extension of Head Start programs to follow through with educational efforts for economically disadvantaged children. Project Follow Through examined nine different models of instruction, but Becker and Carnine (1980) restricted their review to the Direct Instruction model. Over 15,000 kindergarten students took part in Project Follow Through in both experimental and control groups. Students in the Direct Instruction model groups clearly demonstrated greater gains in basic academic goals and affective outcomes overall (Becker & Carnine, 1980). However, effective gains were not made in every area, which may not be model program effects but rather, as Becker and Carnine (1980) believe problems with implementation. In general, it has been found that when the intervention model is complex, treatment integrity is low.

Currently, there is no standardized way of measuring the complexity of an intervention and one is clearly needed. Hunter (1982) developed specific guidelines outlined within a Direct Instruction model that includes seven elements of effective instruction and should be implemented in the following order: 1) identifying objectives; 2) knowledge of standards; 3) anticipatory set or hook; 4) teaching, which includes input, modeling, and checking for understanding; 5) guided practice; 6) closure; and 7) independent practice. Although this model was developed as a guide in designing general classroom lessons, these steps can certainly be adapted to small-group or one-to-one instruction as well. In the next paragraph, each element of effective instruction is described in more detail.

In identifying objectives, it is expected that the teacher will inform the student what he/she should be able to do, understand, and find important as a result of a specific lesson. The teacher must then explain the type of lesson to be given, procedures to be followed, behavioral expectations related to the instruction, and what knowledge or skills should be demonstrated. For the anticipatory set, the teacher aims to gain the student's attention by relating the student's experiences to the objectives of the lesson through the activation of background knowledge and/or prerequisite skills. Next, the teacher provides the necessary information to the student for the lesson. This can be done through lectures, films, tapes, videos, pictures or any other means. Modeling is then provided by showing the student examples of the expected end product of his/her work, explaining the critical aspects through labeling, categorizing, and comparing. The student is then taken to the application level, where he/she completes problem-solving, comparison, and summarizing activities. Before moving on from this step, the teacher

must determine whether the student understands the new material through either formal or informal assessment. Once the "teaching" element has been completed, the student is given the opportunity to demonstrate his/her newly learned skill(s) through an activity under the teacher's direct supervision. The teacher then reviews the key points of the lesson and clarifies any misunderstanding. Finally, activities are provided for independent practice to reinforce the newly learned material. The reinforcement practice should be provided in several different contexts to promote generalization of learning. Such an approach may serve as a useful guide for assessing the level or complexity of a treatment.

Resources: Time, Materials, and Agents

The time required to implement treatments has also been found to be a factor related to treatment acceptability by intervention agents, and is related to the level of effective treatment implementation. Happe (1982) surveyed Iowa school psychologists and found that the majority (87%) reported lack of time as the reason for not implementing a plan as intended. Not surprisingly, the amount of time is related to treatment complexity and there is likely to be an interaction between these two factors. The more complex a treatment the more likely it is that it will take more time to implement.

Similarly, the more materials and resources required for an intervention, the lower the level of effective treatment implementation. This is especially likely when teachers are not familiar with the materials and resources required for the intervention and go beyond the scope of their usual classroom instruction (Gresham, 1989). Interventions requiring more than one treatment agent are also likely to be related to a decrease in

effective treatment implementation, as they further complicate the treatment by introducing another variable (i.e. communication between agents), thus allowing for another area of breakdown within the intervention.

Prior to RTI, when a student experienced difficulty with learning and was not demonstrating response to the curriculum, teachers would refer the student for a special education evaluation. Since the change in special education law, teachers must now attempt to intervene in the classroom and monitor student progress prior to referral. Teachers have needed to make changes in their instruction and the way they approach students unresponsive to the core curriculum. Package intervention programs were typically used solely by special education teachers in a pull-out model, so most general education teachers have little to no experience of these types of material. Although research has found the use of materials typically found in classrooms to be related to high treatment integrity, packaged intervention programs must be used to implement interventions to struggling learners. Teacher-made interventions, although they may yield positive results for some students, are not adequate according to research standards. As referenced previously, IDEA (2004) and RTI both require the use of scientific, research-based interventions, which are typically scripted, packaged programs. The reason for this is that, in order to measure its effectiveness, a program must be systematic in implementation. If it is not systematic in implementation, then treatment integrity cannot be measured. Treatment integrity aside, a teacher-made intervention has no evidence to support its effectiveness. Without the support of research, one cannot determine whether the student is not making progress due to an underlying problem or an

ineffective intervention. This decision can only be made when students are provided with evidence-based interventions implemented with integrity.

Perceived and Actual Effectiveness

Perceived and actual effectiveness have been addressed in much of the literature showing that interventions thought to be effective by the intervention agent will be implemented with greater integrity. It has been suggested that treatments showing behavior change more quickly (i.e. providing faster results) have higher levels of treatment integrity. Treatment integrity is actually reinforced by the behavior change (Gresham, 1989) and it is easy to see how an effective intervention can further increase the motivation of the teacher to continue with proper implementation. Teachers with low motivation for treatment implementation (i.e. teachers who expect the child to be removed from the classroom rather than gaining the role of intervention agent) are associated with lower levels of treatment integrity (Ysseldyke, Christenson, Pianta, & Algozzine, 1983). In a study conducted by Shinn, Powell-Smith, Good, and Baker (1997), the relationship between feedback on effectiveness provided to teachers and student outcomes was examined. Twenty-three special education students were reintegrated into the general education classroom and progress continued to be monitored weekly using Reading CBM. At week four, the students' reading rates had not significantly improved, and this information was shared with the teacher in the form of a graph. Four weeks later, progress was calculated again; the students' reading rates had improved significantly and continued to improve through the end of the 16-week study. The increase in reading rate may have been related to teacher awareness of the student's unsatisfactory progress, although these results are limited due to possible historical

effects. Although these factors are presented in the literature as being related to treatment integrity, there is still little known about the condition under which intervention agents implement interventions with integrity (Noell & Gresham, 1993).

Treatment Acceptability

The acceptability of intervention plans or treatments, regardless of intervention effectiveness, has been shown to be another important factor in determining whether or not the intervention agent actually implements the intervention (Eckert & Hintze, 2000; Elliott et al., 1991). Treatment acceptability refers to views and opinions about the treatment held by those responsible for implementing the treatment. In Kazdin's (1980) definition of treatment acceptability, the variables dependent upon acceptability are the appropriateness of the treatment for the problem, the fairness of the treatment, the reasonability of the treatment, the intrusiveness of the treatment, and whether the treatment meets with conventional notions about what treatment should be. Logically, it would seem that intervention agents who find a treatment acceptable (i.e. appropriate, fair, reasonable, unobtrusive, and in accordance with their own notions about treatment) are more likely to follow through with the implementation. Wolf (1978) provides a review of the importance of social validity in conducting research, drawing on an example from a study conducted by Braukmann, Kirigin, and Wolf (1976) that used the Achievement Place model. The Achievement Place model was a community-based, family, behavioral treatment program for young people. During this investigation, Wolf's (1978) understanding of the importance of social validity ratings became clear when a community dismissed the researchers implementing the program. When the Achievement Place model was presented to another community after changes had been

made based on community feedback (Braukmann et al., 1976), positive social validity ratings were demonstrated, which in turn led to the adoption of the model in the community. Wolf (1978) advocates for the development of measurement systems to gather social validity information from consumers of interventions, as interventions or programs with higher social validity or acceptability are more likely to be adopted.

A model of treatment acceptability introduced by Witt and Elliott (1985) explores the reciprocal and sequential relationships among four elements: treatment acceptability, treatment integrity, treatment effectiveness, and treatment use, with treatment acceptability at the top. If there is treatment acceptability, treatment integrity will be high and if treatment integrity is high, the probability of behavior change will increase. If the behavior is changed, then the acceptability of the treatment by the intervention agent will also increase and therefore be an acceptable treatment for future use. A possible conceptualization the model proposed by Witt and Elliott (1985) is demonstrated in Figure 1. Elliott et al. (1991) indicate a relationship between treatment acceptability and treatment use, demonstrating that interventions with high acceptability lead to an increase in the use of the intervention. A relationship between treatment integrity and behavior change has also been supported (Gresham, 1989), demonstrating that high levels of treatment integrity are related to increased behavior change.

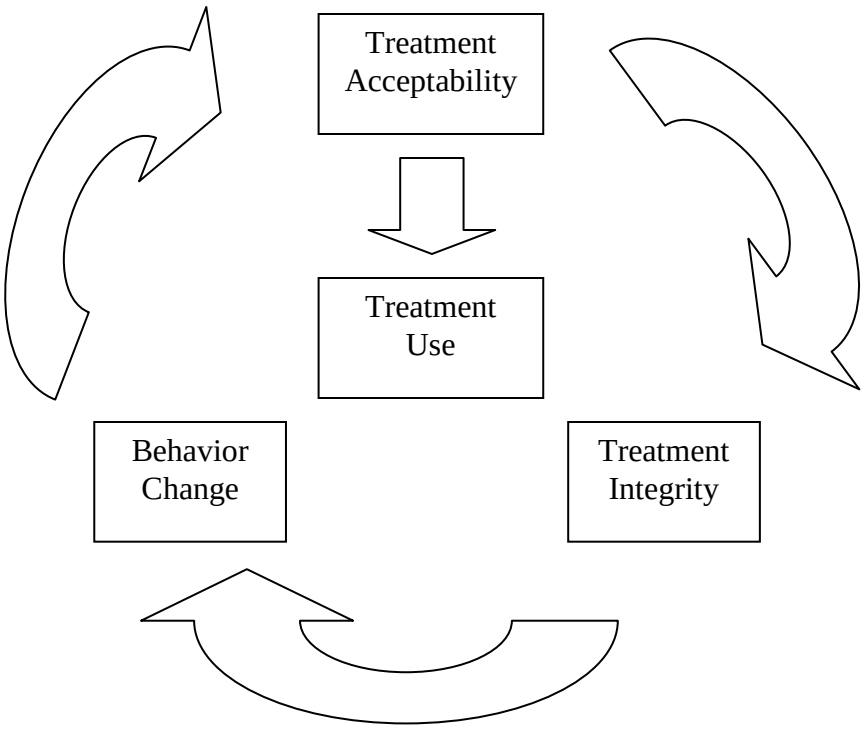


Figure 1: Conceptualization of Witt and Elliott's (1985) proposed model of treatment acceptability.

A similar model developed by Reimers, Wacker, and Koeppl, (1987) added a knowledge component to Witt and Elliott's (1985) model arguing that a treatment must first be understood before the level of acceptability can be assessed. Poor understanding of an intervention leads to low implementation integrity, which in turn leads to low effectiveness (behavior change). Good understanding may lead to either low acceptability or high acceptability.

Additional factors have been shown to influence treatment acceptability, including type of language used to describe interventions, consultant involvement, type of treatment, time required for implementation, problem severity, and intervention agent variables such as knowledge of behavior principles and years of experience. In a study conducted by Witt, Moe, Gutkin, and Andrews (1984), the relationship between the types

of language used to describe interventions and teacher ratings of acceptability of the intervention was investigated. The intervention was described either in behavioral terms, in pragmatic terms, or in humanistic terms. The findings indicated that when the intervention was described in pragmatic terms as opposed to behavioral or humanistic terms, acceptability ratings of the intervention were high. Algozine, Ysseldyke, Christianson, and Thurlow (1983) investigated the relationship between treatment acceptability and teachers' intervention preferences, depending on the level of consultation provided. Case studies of 40 treatment alternatives applied to three types of classroom problem were presented to teachers. Teachers gave a higher acceptability rating to interventions that could be implemented directly than to those that depended on a consultation with a psychologist.

Witt, Elliott, and Martens (1984) investigated the relationship between type of treatment and treatment acceptability. A total of 180 pre-service teachers were asked to rate acceptability of six treatments on the Intervention Rating Profile-20. Six treatments were provided for changing target behaviors of daydreaming, obscene language, and destroying property: three positive (praise, home reinforcement, and token economy) and three negative (ignoring, response cost, and seclusion time-out). Pre-service teachers rated positive treatments more acceptable than negative treatments for the same problem. Elliott, Witt, Galvin, and Peterson (1984) replicated the Witt, Elliott, et al. (1984) study using the same six interventions and again found positive treatments to be rated more acceptable. In addition, an interaction effect between treatment type and problem severity was found. Severity has been operationalized as both the degree of intensity of inappropriate behavior and the number of children who demonstrate inappropriate

behavior (Elliott et al., 1991). Elliott et al. (1984) operationalized the definition of severity as the degree of intensity. Teachers were asked to read several case descriptions of a student whose misbehaviors were low (daydreaming), moderate (obscene language), or severe (destruction of property) and rate the acceptability of one of three possible positively oriented interventions that ranged in complexity. The three possible interventions were praise (low complexity), home reinforcement (moderate complexity), and a token economy (high complexity). The results showed that the least complex intervention (praise) was rated as the most acceptable treatment for the least severe problem (daydreaming), while the most complex intervention (token economy) was rated as the most acceptable treatment for the most severe behavior problem (destroying property). Teacher ratings of the acceptability of reductive interventions ranging in complexity included ignoring (low), response cost lottery (moderate), and seclusion time-out (high). As with the positively oriented interventions, the results showed that the least complex intervention (ignoring) was the most acceptable treatment for the least severe behavior problem (daydreaming) and the most complex intervention (seclusion time-out) was the most acceptable treatment for the most severe behavior problem (destroying property). When the severity of the problem was high, teachers rated more complex interventions as acceptable.

The study conducted by Witt et al. (1984) also found problem severity to influence teacher rating on treatment acceptability. The study explored the relationship between several variables: teacher time involvement; intervention type; and severity of the behavior problem. Teachers evaluated treatments on the Intervention Rating Profile-20 after reading descriptions of each intervention, including the amount of time required

for implementation. Time required for treatment implementation ranged from low (less than 30 minutes per day) to high (more than three hours per day, including maintenance). Teachers preferred interventions that were more time-efficient, but when the presenting problem was rated as more severe, expectations about time were adjusted. These findings provide support for severity of the problem as a salient factor for implementation integrity and provide useful information for practitioners in designing intervention plans.

According to a study conducted by McKee (1984), teacher knowledge of behavior techniques was found to be related to treatment acceptability. Teacher knowledge was assessed using a modified 16-item version of the Knowledge of Behavior Principles as Applied to Children Test (KBPAC) (O'Dell, Tarler-Benlolo, & Flynn, 1979). Based on the scores of the assessment, teachers were assigned to either a high-knowledge or low-knowledge group using a median split technique. Teachers in the high-knowledge group rated treatments as more acceptable than those in the low-knowledge group. Finding from this study give support to the understanding component in the Reimers et al. (1987) model mentioned previously. Teachers demonstrating a greater understanding or knowledge of a treatment rated treatments as more acceptable.

An inverse relationship has been found between teachers' experience (i.e. number of years taught) and treatment acceptability (Witt, Moe, et al., 1984). The study by Witt, Moe, et al. (1984) is described in the previous section where the relationship between type of language used to describe treatments and treatment acceptability is examined. Teachers having more years of experience rated acceptability of all treatments as low.

When treatment acceptability by teachers (or intervention agents) is high, treatment integrity of interventions has also been found to be high. From the literature,

we also know which factors are related to high acceptability of treatment by teachers. Type of language used to describe interventions, consultant involvement, type of treatment, time required for implementation, problem severity, and intervention agent variables such as knowledge of behavior principles and years of experience are all factors related to teacher acceptability. These factors deserve more attention in research and in practice when designing interventions for teacher implementation. Additionally, researchers and practitioners should also be aware of teacher variables, such as years of experience, knowledge of behavioral principles, type of training and classroom management techniques used. If interventions can be created with these factors in mind, treatment acceptability by teachers and treatment integrity are likely to increase.

Assessment of Treatment Integrity

Treatment integrity has been assessed using both direct measures that are widely used within school settings (i.e. naturalistic and systematic observations) and indirect measures (i.e. rating scales, self-monitoring, self-report, and behavioral interviews). Each type of measurement system has both advantages and disadvantages (Gresham, 1996). Direct and indirect measures are discussed and critiqued in the following section.

Direct Measures

Typically, when treatment integrity is measured directly, an observer has a list of implementation steps that are indicated as being present or not present during the implementation of the intervention. In reporting the integrity of a particular session, the number of steps completed is divided by the total number of steps to yield a percentage. Several researchers make suggestions about how to assess intervention integrity (Dane & Schneider, 1998; Gresham, 1989, 1996; Gresham, MacMillan, Beebe-Frankenberger, &

Bovian, 2000; Lane, Bocian, MacMillan, & Gresham, 2004). Three steps are typically involved in designing a direct observation system for the assessment of treatment integrity: 1) clearly specify all the treatment components in operational terms; 2) record the occurrence and non-occurrence of each treatment component; and 3) calculate the percentage of treatment components implemented by the intervention agent (Gresham, 1989,1996; Lane et al., 2004). For the third step of the system, Lane et al. (2004) add that both component and daily or session integrity should be computed. Component integrity is the average rating for a specific intervention component and measures how much of the intervention was implemented, while daily or session integrity is the average rating for all intervention components combined and measures how well the intervention was measured (Gresham et al., 2000). Component and daily integrity can be further categorized according to a review of treatment integrity in prevention research conducted by Dane and Schneider (1998). Component integrity, which looks at the content of the intervention, can be further categorized by treatment adherence, exposure/dosage, and program differentiation. Daily integrity, which looks at the process of the intervention, can be further categorized by the quality of delivery and participant responsiveness (Dane & Schneider, 1998). Intervention adherence measures the extent to which the intervention agent follows the specific objectives of the intervention, which is what many researchers have focused on. Exposure or dosage, which has received very little attention in research studies, measures the number, length and frequency of intervention occurrences. Program differentiation measures the different components of the intervention to distinguish between similar interventions and ensure only relevant components are being used. This is an area that many researchers believe is a necessary

focus for future research in treatment integrity (e.g. Gansle & Noell, 2007). Quality of delivery measures the more qualitative aspects of the intervention, focusing on effective qualities of the intervention agent. Finally, participant responsiveness measures the engagement level of the student. Research investigating the last two elements of treatment integrity is scarce. Dane and Schneider (1998) point out the lack of consistency in the research when measuring intervention integrity and suggest focusing on these five elements of integrity. By doing so, there will be more consistency across the board in integrity research, making it easier to draw conclusions.

In using direct forms of assessment, although it is still undecided in the literature how many or how long observations should be in order to accurately sample true behavior, it has been suggested that 3–5 observations each of between 20 and 30 minutes is acceptable as a rough guideline (Gresham, 1996). Although direct observation of behavior is ideal, it is not guaranteed that the observation session actually captures typical behavior. In collecting direct observation data, the research runs the risk of reactivity. Reactivity of observations is when those being observed change behavior due to the presence of an observer (Gresham, 1996; Hintze & Matthews, 2004; Merrell, 2003). It is likely that the treatment agent will follow the treatment protocol more strictly while being observed.

Indirect Measures

Types of indirect measurement, which are less obtrusive, include self-monitoring, rating scales, and interviews. For the most part, these indirect measurements are developed by researchers to fit particular research questions. Collection of self-monitoring data relies on the intervention agent to be truthful and accurate in the

reporting of treatment integrity. Hagermoser Sanetti and Kratochwill (2009) found self-report data to be more accurate when collected more frequently: on a daily basis as opposed to a weekly or monthly basis as reported in previous research (e.g. Noell & Gansle, 2006). More research is necessary to explore this finding, but promising support has been provided for the use of self-report data as an adequate measurement tool for treatment integrity.

Rating scales that have been used to assess treatment acceptability in the literature include: the Behavior Intervention Rating Scale (BIRS) (Von Brock, 1985; Von Brock & Elliott, 1987); the Children's Intervention Rating Profile (CIRP) (Witt & Elliott, 1985); the Intervention Rating Profile-20 (IRP-20) (Witt & Martens, 1983); the Intervention Rating Profile-15 (IRP-15) (Martens, Witt, Elliott, & Darveaux, 1985); and the Treatment Evaluation Inventory (TEI) (Kazdin, 1980). The BIRS was found to have an internal reliability ranging from 0.87 to 0.97 and a criterion validity ranging from 0.52 to 0.78 with the Evaluative Scale of the Semantic Differential (Elliott & Treuting, 1991). The CIRP was found to have an internal reliability ranging from 0.75 to 0.89 (Martens & Meller, 1989) and a criterion validity ranging from 0.67 to 0.69 with the Treatment Expectancy Scale (Waas & Anderson, 1991). The IRP-20 was found to have an internal reliability of 0.89 (Witt & Martens, 1983) and a criterion validity of 0.86 with the Evaluative Scale of the Semantic Differential (Witt, Elliott, et al., 1984). The IRP-15 was found to have an internal reliability of 0.98 and a criterion validity of 0.86 with the Evaluative Scale of the Semantic Differential (Martens, Witt, Elliott, & Darveaux, 1985). The TEI was found to have an internal reliability of 0.89 (Kelly, Heffer, Gresham, & Elliott, 1989).

Unfortunately, indirect methods, which depend on reporting from the intervention agent, can reflect bias and possibly not provide an accurate depiction of the actual behavior. There are several types of response bias that may occur in self-report measures: acquiescence, social desirability, faking, and deviation (Merrell, 2003). Acquiescence is the tendency for self-reporters to consistently choose answers in the same direction (i.e. answering yes to all questions). In the case of social desirability, self-reporters will choose answers that are socially acceptable rather than their true feelings. Faking is similar to social desirability, but in this case the self-reporter will choose answers that create a positive impression. Finally, deviation refers to an unconventional or unusual way of answering the questions. For the types of self-report measures used for assessing treatment integrity, faking is most likely to be a problem since intervention agents are likely to want to give the researcher the impression that they are implementing the intervention as intended. Owing to the disadvantages of both direct and indirect assessment methods, it is recommended that when conducting treatment integrity research, several methods and sources are used to collect data (Gansle & Noell, 2007; Gresham, 1996).

Summary

Purpose of the Current Study

In reviewing the literature, the importance of treatment integrity research is clear in both legal and practical arenas, as are its implications for successful intervention implementation. Ultimately, in an RTI approach, "decisions regarding changing or intensifying an intervention are made based on how well or how poorly a student responds to an evidence-based intervention that is implemented with integrity"

(Gresham, 2007, p.10). The key to this statement and the basis for RTI being successful is that the intervention is "implemented with integrity". Without the assurance of treatment integrity, one cannot definitively judge whether the instructional programming designed to improve student performance was truly unsuccessful as a function of student learning disabilities. If schools intend to comply with the most current special education laws, they must take the necessary steps to ensure interventions are being implemented and that they are being implemented with integrity. Treatment integrity practices should no longer be ignored or seen as a separate step in implementing interventions, but should be considered a key component within the intervention itself that is necessary for successful implementation. Ultimately, it is necessary to ensure the integrity of the decisions being made for students (Shinn, 2007).

Although many researchers have begun to investigate technical issues, factors of implementation, and assessment of treatment integrity, there is still very little known for how to ensure successful adherence to an intervention. To further the research base, the direction of integrity research must develop in several ways. First, it is essential to further understand the technical issues related to treatment integrity. A greater emphasis must be applied to the specificity of definitions. Dependent and independent variables must be operationalized to enable a more reliable measurement of variables and for replication of research. Additionally, questions must be answered about the intervention dosage necessary to elicit the intended result. Measurement of treatment exposure and treatment adherence must also be a priority in research, as the current tools of assessment may be inaccurate and unreliable (i.e. self-report data). The crucial components of an intervention must also be identified.

Second, it is essential to identify those crucial factors or steps necessary for the intervention to be successful (Gansle & Noell, 2007). Identification of factors (i.e. complexity of treatments, time required, materials/resources, number of treatment agents, perceived and actual effectiveness of treatments, motivation of treatment agents, severity of the problem, and acceptability of treatments) that predict high treatment integrity and – even more importantly – further predict positive student outcomes must continue to be examined. With the most recent changes to educational law and a solid foundation of treatment integrity research, new research should focus on answering those unanswered questions.

The results of the literature review clearly support factors related to treatment integrity of interventions. However, the research has examined each component separately when exploring the relationship with implementation integrity. Hagermoser Sanetti and Kratochwill (2009) propose that future research focus on whether interaction effects occur between factors related to treatment integrity (moderators) and whether variables identified as related to treatment integrity are also related to treatment outcomes. The current study proposes to further explore the theory behind treatment integrity and the suggestions provided by Hagermoser Sanetti and Kratochwill (2009) by grouping certain related factors together to explain the various levels of treatment integrity in intervention implementation. It is hypothesized that treatment integrity varies as a function of both complexity and acceptability. Subsequently, these primary factors are related to treatment integrity, which in turn is related to student outcomes.

At the present time, there are no tools used to measure the complexity of an intervention. In conducting research for the current study, the researcher identified the

need and explored possible ways to measure the complexity of interventions in a standardized manner. Hunter's (1982) Direct Instruction model for guiding instructional design was used for the development of a tool to measure complexity, with higher scores indicating a closer alignment with Hunter's (1982) model and a higher complexity. A logical argument could be made that using a direction instruction model makes the implementation of the intervention less complex owing to the scripted nature of the model. Although a model that is organized in a prescribed manner may be easier to read and follow, the complexity of delivery makes it more difficult to implement. The model goes beyond simply reading the curriculum or intervention protocol as a script. The Direct Instruction model requires the implementation of several interrelated treatment components, such as preparation of materials, enthusiasm to gain the students' attention, behavioral expectations, assessment of student understanding, and providing opportunities for both guided and independent practice, in addition to the "teaching" of the lesson at a rapid pace of instruction. This model has similar elements to the previously mentioned intervention (i.e. Direct Instruction model) that have been considered to be highly complex according to the literature (Becker & Carnine, 1980).

The Complexity Factor includes four factors previously found to be related to treatment integrity: 1) complexity of the treatment; 2) time required in implementing the treatment; 3) materials and resources required for the treatment; and 4) the number of treatment agents required for treatment implementation. This was decided because each of these individual factors is likely to be related to one another. For example, Gresham (1989) notes an interaction between complexity and time, and that when a treatment is more complex it is very likely that it will take more time to implement. The case can also

be made for the type of materials required and the number of agents required for intervention implementation.

The Acceptability Factor will include three additional factors previously found to be related to treatment integrity: 1) perceived effectiveness of the intervention; 2) severity of the problem; and 3) acceptability of the intervention. These three factors were grouped together because they all measure teacher perceptions. It is likely that the teacher will rate each of these factors in a similar way. If the teacher feels that the intervention is effective and the problem is severe enough to warrant the intervention, then he/she will likely find the treatment to be acceptable. Therefore, it is believed that each of these individual factors discussed in the literature should be combined to form the Acceptability Factor in the present study.

Lastly, a necessary element for the measurement of treatment integrity is a sound technology for directly and indirectly assessing integrity (Gansle & Noell, 2007) that could also be used across treatment agents, situations, and time (Gresham 1989). Additionally, without having generic measurement tools that can be used for a variety of interventions and problems (i.e. academic vs. behavioral), comparisons with other studies are much more difficult. With a universal technology for assessing treatment integrity comparisons of different treatments implemented at various levels of integrity can be measured and will help to understand what treatments produce effects at specified levels of integrity (Gresham, 1989).

Research Questions

The following research questions and hypotheses will be explored in the current study.

1) What are the direct relationships among implementation factors, treatment integrity and reading fluency?
    a. **Hypothesis:** Consistent with past research, it is hypothesized that low scores for the complexity factor will predict a high level of treatment integrity and that high scores for the acceptability factor will predict a high level of treatment integrity. High scores for treatment integrity will predict greater gains in reading fluency.
2) How does treatment integrity mediate these relationships?
    a. **Hypothesis:** It is hypothesized that low scores for the complexity factor will predict greater gains in reading fluency as mediated by treatment integrity and high scores for the acceptability factor will predict greater gains in reading fluency as mediated by treatment integrity.

CHAPTER 3

METHODOLOGY

The purpose of this chapter is to describe the methods used in conducting the research project. Information is provided on the setting, subjects and method of subject recruitment. Lastly, the procedures and the outcome measures are described.

Setting and Participants

Setting

Data were collected during the 2009/2010 school year at an elementary school currently receiving funding from a Reading First grant in the southwestern part of the United States. The district is one of the largest in the country and includes 213 elementary schools, 59 middle schools, 48 high schools, 24 alternative schools, and 8 special schools. The district employs over 38,000 employees that serve over 309,000 students. The following demographic data for the school are reported for the 2009/2010 school year. The ethnicity of the students is 41% Hispanic, 34% Caucasian, 14% African American, 9% Asian/Pacific Islander, and 0.7% Native American. Approximately 20% of students have limited proficiency in English.

The elementary school in which the study was conducted was one of two schools that the researcher was assigned to as a school psychologist intern for the 2009/2010 school year. Participants were not drawn from the second elementary school owing to the lack of a universal screening measure and implementation of Tier-II interventions. Approval from the participating school principal and school district department of research, along with sponsorship from the director of psychological services, was obtained prior to the study.

Participants

The participating elementary school did not meet the standards for Adequate Yearly Progress (AYP) for the 2008/2009 school year and was classified as "In Need of Improvement – Year Two". The following demographic data for the school are reported from the 2008/2009 school year. In comparison to the district, this school had a predominantly Hispanic population (83.6%), with 61% of the students classified as having limited proficiency in English. Other demographic data include: 7% Black/African American, 6% Caucasian, and 2% Asian/Pacific Islander. A total of 92% of the students qualified for a free or reduced-price lunch. There was a 34% transience rate and the student to teacher ratio ranged from 16:1 to 27:1, with smaller class sizes in the kindergarten and first-grade classes.

Universal Tier I Screening was conducted in September 2009. As a result, a total of 122 student participants in grades 2 through 5 were placed into Tier II interventions for reading problems based on their screening score and were eligible for participation. All general education students with scores below benchmark criteria on the DIBELS Oral Reading Fluency (DORF) test who gave their consent and whose parents gave their consent to take part in the study were included. Students receiving special education services were excluded from the study, owing to additional reading interventions and instruction time received by special education teachers. Therefore, students participating in the study were receiving only Tier I core instruction and either the Tier II Voyager Passport (Voyager Expanded Learning, Inc., 1994) intervention or the Tier II Read Well (Voyager Expanded Learning, Inc., 1984) intervention. In total, 31 second-grade, 17 third-grade, 34 fourth-grade, and 40 fifth-grade students participated, of whom 56 were

girls and 66 were boys. Ethnicity of the sample included 86% Hispanic, 7% Caucasian, and 6% African-American. Seventy-four percent of participants were classified as English Language Learners. Twenty teachers serving as intervention agents participated, 19 of whom were female. Years of teaching experience ranged from one year to 34 years, with 10 being the median number of years taught. Three teachers and 15 students were excluded from data analysis owing to missing data.

Procedure

The investigation was conducted between October and November 2009 during the fall semester of the school year. The Acceptability of Intervention Measure (AIM) pre-survey was completed by participating teachers prior to the start of interventions. The researcher trained five school psychologist interns on the Intervention Integrity Checklist observation tool to assist with data collection throughout the seven-week intervention period. Descriptions of the interventions to be observed and the observation tool were provided to the observers and any questions were addressed prior to the observations. During the intervention period, each intervention group was observed a total of three times. The three observation days for each group were randomly assigned by week. An average of the three observations was computed to obtain a score of treatment integrity. Observations were completed on Thursday and Friday of each week unless the school was closed due to a holiday or staff development day, in which case observations were completed on Wednesday. Also, owing to federal and state holidays, a class party (i.e. Halloween celebration) and staff development days, interventions were implemented on only 31 out of the possible 35 days of the intervention period. After the intervention

period, teachers completed the Factors of Implementation – Self Report Measure and the AIM post-survey.

Two significant changes to the original methods proposal needed to be made. First, a nine-week intervention period was originally foreseen but, owing to the 12-month school schedule, adjustments in research design needed to be made. The entire school operates on a "track" schedule, which means that the students were out of school for a five-week period beginning at week eight of the intervention period. It was therefore decided to shorten the intervention period to seven weeks instead of having a five-week break with no intervention during those five weeks. Shortening the intervention period certainly lessened the number of days the intervention was implemented and the time available to allow growth in reading. Second, it was originally anticipated that a variety of reading interventions would be implemented. However, owing to the large number of students in need of intervention, the participating school followed a standard protocol model that included only two reading interventions. Also, owing to the timing of the present study, it was not possible to include Tier III reading interventions. These changes in methodology and their implications are covered more thoroughly in the discussion section.

Response to Intervention – District Model

During the 2004/2005 school year, 11 schools within the district began piloting an RTI model. A team of school psychologists formed to train individual schools on the core principles of RTI (full-day PowerPoint training) and the district procedures necessary for implementation (half-day grade level training). Since the implementation of RTI in the pilot schools in the 2004/2005 school year, additional schools have been

trained each year up to the present. RTI is now a district-mandated practice for identifying children with a specific learning disability, although individual schools vary in the level of RTI implementation that has been reached.

The RTI model, adopted district-wide, consists of three tiers. Tiers I (core instruction) and III (intensive instruction of a targeted skill) are consistent with the literature, whereas Tier II varies. The district breaks Tier II down into two levels based on benchmark assessments: Level I consists of students who fall into the strategic (minor delays) category, while Level II consists of students who fall into the intensive (major delays) category. Progress is monitored every other week for students at the strategic level and weekly for students at the intensive level. A problem-solving team was in place at each school to develop intervention plans and make decisions based on student progress. The membership of the problem-solving team varied depending upon the school. The benchmarking system in place at each school also varies depending upon resources. Aimsweb (Pearson Education, Inc., 2008), DIBELS (Dynamic Measurement Group, Inc., 2009), and mCLASS (Wireless Generation, 2010) are the most common benchmarking systems used throughout the district. There is no psychometrically sound assessment system in place for the measurement of treatment integrity at any level of instruction or intervention practice. Although the treatment integrity of interventions was not typically assessed, certain safeguards were put in place to facilitate procedural integrity. An intervention assessment protocol (see Appendix A) developed by the district is completed by the multidisciplinary team once a suspicion of a disability is identified, and a referral is made to the multidisciplinary team for a special education evaluation. Accountability of procedural integrity is delegated to classroom teachers.

In the development of a Tier III intervention plan, three elements are necessary for implementation: 1) Intervention Plan (see Appendix B); 2) Intervention Log (see Appendix C); and 3) Intervention Graph (see Appendix D). Additionally, each teacher required to collect DORF data has a palm pilot for data collection. Data from the palm pilot are uploaded to the mCLASS (Wireless Generation, 2010) website weekly to ensure the procedural integrity of data collection for monitoring progress.

The Tier III intervention plan was developed by the individual grade level's problem-solving team, which consists of all grade level teachers (i.e. each grade level is its own team) and occasionally a specialist, depending on the area of concern. The individual classroom teacher used the Tier III intervention plan to carry out the intervention. The intervention log was used to document each day the intervention was actually implemented and the amount of time that the student received the intended intervention. Progress was monitored weekly and plotted on the intervention graph. If a student did not make adequate progress after the nine weeks in response to the Tier III intervention (18 weeks of data in total, including the data collected during Tier II interventions), the student was referred for a special education evaluation.

Response to Intervention – Participating School

The participating school first implemented RTI as a system of service delivery for the current school year (2009/2010). The adopted model consisted of a combination of a standard protocol approach and a problem-solving model within a three-tier framework. Tier I consisted of Harcourt Trophies (Harcourt, Inc., 1919) as the core reading curriculum for grades 1–5. This was a new core reading program for grades 1, 2, and 3, although it had been used in grades 4 and 5 for four previous years. The reading block

was 90 minutes long and consisted of whole-group instruction as well as differentiated small-group instruction. Benchmark assessments in the area of reading occurred three times per year (September, February, and June). Subsequent to the September benchmark, students whose performance was below grade level benchmark were moved into either Tier II – Level I or Tier II – Level II intervention groups. Each grade level had a 40-minute block built into its daily schedule specifically for Tier II interventions. Students who scored at or above the DIBELS (Dynamic Measurement Group, Inc., 2009) benchmark criteria for minimal competency were considered to be "low risk" and were given independent work at their level to complete during intervention time. Students who scored below the benchmark criteria for minimal competency, but not so low as to be considered "at risk", were considered to be at "some risk" and received Voyager Passport Intervention (Voyager Expanded Learning, Inc., 1994) and their progress was monitored every other week. Students whose performance were lowest and were considered to be "at-risk" received the Read Well Intervention (Voyager Expanded Learning, Inc., 1994) and were their progress was monitored weekly.

This model is fluid, meaning that students move between the levels based upon their performance on the DIBELS (Dynamic Measurement Group, Inc., 2009) assessment. The following procedures took place after the conclusion of the present study, but are described here to provide a clearer understanding of the RTI model implemented within the participating school. Progress monitoring data were assessed by the lead RTI team after nine weeks of intervention. Students who demonstrated adequate progress in the Tier-II – Level I intervention groups and achieved three consecutive scores in the "low risk" category were moved out of the intervention group. Students

who demonstrated adequate progress in the Tier-II – Level II intervention groups and achieved three consecutive scores in the "some risk" category were moved from the Read Well intervention group to a Voyager Passport (Voyager Expanded Learning, Inc., 1994) intervention group. Students who demonstrated inadequate progress within Tier II – Level I were moved to a Tier II – Level II intervention group, while students who demonstrated inadequate progress within Tier II – Level II were referred to the grade level problem-solving team. The team then developed an intervention plan more specific to the student's problem (Tier III), which was implemented in addition to the core reading (Tier I) and reading intervention times (Tier II). Procedures outlined previously for Tier III interventions were followed. Treatment integrity of Tier III interventions was not assessed, as they were implemented after the conclusion of the present study.

Independent Variables

Tier II Interventions

Reading interventions selected for Tier II were purchased by the literacy specialist at the end of the previous school year (2008/2009) prior to the researcher's assignment to the participating elementary school. Voyager Passport (Voyager Expanded Learning, Inc., 1994) is a comprehensive, research-based intervention designed for students in grades K through 5 who are struggling to read. It has shown the most success for students who are approximately one to two years behind in reading and was therefore implemented at the strategic level. It contains systematic and explicit lessons that integrate the five essential components of reading (phonemic awareness, phonics, fluency, comprehension, and vocabulary) and should be delivered through the standard protocol provided. Voyager Passport (Voyager Expanded Learning, Inc., 1994) was

designed for implementation in a small group of up to six students to supplement the core reading curriculum. Two components make up the Voyager Passport (Voyager Expanded Learning, Inc., 1994) daily curriculum: 1) word works, which provides instruction in phonemic awareness, letter–sound recognition, word reading, and sight words; and 2) read to understand, which provides students the opportunity to practice the newly learned skills with instruction for vocabulary building and comprehension. A fluency book is also included for additional practice. Voyager Passport (Voyager Expanded Learning, Inc., 1994) is based on five steps. The first step includes teaching, modeling, and probing. Material is explicitly taught to the student and the purpose for learning each concept is presented. Each concept and skill is modeled for the student, followed by probing to ensure initial understanding of what was taught. The second step is guided practice, in which students are presented with short tasks related to the concept. Each student's success is closely monitored and immediate corrective feedback is provided when necessary. Student understanding is ensured before moving on to independent practice, which is the third step in the model. Students are provided opportunities to practice the newly learned skill independently. The fourth step is the cumulative review, where the previously taught elements are systematically reviewed. New concepts are integrated into previously taught concepts to provide students with continual practice and reinforcement. The fifth and final step in the model is assessment of each concept. Teachers have the opportunity to informally assess student responses for accuracy throughout each lesson. In addition, every fifth lesson includes a formal assessment (Adventure Checkpoint) to measure students' understanding and a skill review is provided as needed. Based on the results of the assessment, teachers either

move forward with instruction or re-teach skills as necessary. Students are also encouraged to share what they have learned with their families after each Adventure Checkpoint is completed.

Read Well (Voyager Expanded Learning, Inc., 1994) is a research-based reading curriculum for the primary grades (K–2), consisting of whole- and small-group lessons. Since the program was used as an intervention rather than as a core curriculum, whole-group lessons were also taught in the small-group format. Read Well (Voyager Expanded Learning, Inc., 1994) contains systematic and explicit lessons that integrate oral language and the five essential components of reading (phonemic awareness, phonics, fluency, comprehension, and vocabulary) and should be delivered through the standard protocol provided. Daily lesson plans consist of 30 minutes and include two components: decoding practice and story reading. Teacher materials include a teacher's guide, sound and word cards, smooth and bumpy blending cards, spring toys (used for stretch and shrinking of words), and an assessment manual. Each student is provided with a decoding book, a story book, a comprehension and skill work book, and homework. Daily lessons begin with a review of the previous lesson. Using the decoding book, the new sound is introduced, practice with the new sound is provided, spring toys are used to stretch and shrink words with the new sound (e.g. cats, caaatsss, cats), sounding out practice is provided, and a list of sight words are practiced. Once decoding practice is complete, story reading begins. The story is introduced and relevant vocabulary is reviewed and practiced. Read Well (Voyager Expanded Learning, Inc., 1994) provides the choice of following a three-, four-, six-, or ten-day lesson plan; the ten-day lesson plan was implemented at the participating elementary school, providing

students with additional review and practice for each story. Homework practice was also provided each day.

Interventions were implemented daily during each grade level's 40-minute intervention time block. Intervention time blocks for each grade level were as follows: second grade – 9:10 to 9:50 a.m., third grade – 11:45 a.m. to 12:25 p.m., fourth grade – 8:15 to 8:55 a.m., and fifth grade – 10:10 to 10:50 a.m. Intervention groups were formed by the literacy specialist based on student DORF scores to keep variability of reading ability low. In general, classroom teachers were responsible for implementing the Voyager Passport (Voyager Expanded Learning, Inc., 1994) intervention, while instructional aides, a special education teacher, and a literacy specialist were responsible for implementing the Read Well (Voyager Expanded Learning, Inc., 1994) intervention. However, two second-grade classroom teachers were asked by the literacy specialist to implement the Read Well (Voyager Expanded Learning, Inc., 1994) intervention instead, owing to a larger number of students in second grade requiring the more intensive intervention. Intervention agents were responsible for monitoring progress for each student in their intervention group.

Seven factors of intervention implementation found to be related to treatment integrity were measured and categorized into two groups: 1) Complexity factors and 2) Acceptability of Intervention factors. The Complexity factors that were assessed by the Factors of Implementation Checklist included complexity of treatments, time required in implementing treatments, materials/resources required for treatments, and number of treatment agents required. The Acceptability of Intervention factors assessed by the AIM included perceived effectiveness of treatments, severity of the problem, and acceptability

of the treatment by the treatment agent. In the following section, each of the seven factors is operationally defined.

Factors of Implementation

Complexity Factors

- Complexity of the treatment was defined by the following elements of effective instruction, as proposed by Hunter (1982): 1) identifying objectives; 2) knowledge of standards; 3) anticipatory set; 4) teaching, which includes input, modeling, and checking for understanding; 5) guided practice; 6) closure; and 7) independent practice. These elements were used to develop a treatment integrity protocol (Factors of Implementation Checklist) for the purposes of assessing the complexity of reading fluency interventions used in the study. Operational definitions of each element are included in the checklist.
- Time required in implementing treatments was the total time that the intervention agents spent in three areas: 1) preparing for the intervention, including photocopying, gathering materials, practicing with the materials, and organizational planning for the intervention; 2) implementing the intervention, which is from the time the intervention began to when it ended; and 3) collection of progress monitoring data, which included organizing and printing materials, data collection, and recording the student's progress.
- Materials/resources required for treatments were considered to be anything that the teacher used outside of his/her regular classroom and was not typically used as part of the general curriculum teaching (i.e. used with the entire class).

- Number of treatment agents required was defined as any person who participated in administering the intended intervention to the student. This included any person who participated in collecting relevant materials, implementing the actual intervention, or monitoring student progress.

Acceptability of Intervention Factors

- Perceived effectiveness of treatments was defined as the score received on the treatment effectiveness scale as indicated by the intervention agent. Scores ranged from 1 (not at all effective) to 6 (very effective).
- Severity of the problem was defined as the score received on the severity question on the AIM as indicated by the intervention agent. Scores ranged from 1 (not at all severe) to 6 (very severe).
- Acceptability of treatment was defined as the score received on the AIM.

Measures

Acceptability of Intervention Measure (AIM)

The AIM was adapted from the Syracuse School-Based Intervention Team (SBIT) Introductory Staff Survey and Referring Teacher Acceptability Questionnaire (Wright, 2008), and the Behavior Intervention Rating Scale (BIRS) (Von Brock & Elliott, 1987). Several questions on the SBIT were taken directly from the BIRS. The BIRS included questions seeking teacher perceptions on behavior rather than academic interventions. Statements regarding behavior were not applicable for the purposes of the current study and therefore either they were not included or wording was changed to reflect reading. Examples of statements that were excluded in the development of the AIM include: 1) "The intervention would improve the child's behavior to the point that it would not

noticeably deviate from other classmates' behavior"; and 2) "Soon after using the intervention, the teacher would notice a positive change in the problem behavior". The following statement from the BIRS (Von Brock & Elliott, 1987), "The intervention is a fair way to handle the child's problem behavior" was changed to "The intervention is a fair way to handle the target problem". Several items taken directly from the BIRS included: "I liked the procedures used in this intervention", "This intervention was a good way to handle the problem", and "Overall, the intervention would be beneficial to the child". Questions adapted from the SBIT (Wright, 2008) assessed teacher perceptions on severity of the problem and time/resources available to implement the intervention. The AIM was a 13-question survey developed to assess the Acceptability factor, which includes questions relating to acceptability of the intervention, perceived effectiveness of the intervention, and severity of the problem and addresses teacher perceptions. Questions are answered on a six-point scale ranging from "Strongly disagree" to "Strongly agree". Teachers completed the AIM survey pre-intervention (see Appendix E) and post-intervention (see Appendix G). Slight changes were made on the post-intervention survey to reflect the past tense. Reliability coefficients for both the pre and post AIM surveys were established using Cronbach's Alpha and found to be at an acceptable level. The pre AIM survey yielded a reliability coefficient of .913 and the post AIM survey yielded a .912.

Intervention Integrity Checklist – Observations

The Intervention Integrity Checklist (see Appendix F) was used to document and record treatment integrity. As a broad measure of treatment integrity, the Intervention Integrity Checklist is adaptable to a variety of interventions and comprised 13 questions

that assess the extent to which each component of the intervention was implemented. The Intervention Integrity Checklist was developed by the researcher based on Hunter's (1982) model of direct instruction. Each component was measured on a six-point scale ranging from "Not at all Implemented" to "Implemented Completely." Observations using the Intervention Integrity Checklist were completed on three randomly selected occasions for each intervention group. A reliability coefficient for the Intervention Integrity Checklist was established using Cronbach's Alpha, but was not found to be at an acceptable level (.632). However, in a recent commentary written by Gresham (2009), the recommendation for determining reliability of integrity measures through internal consistency indices is questioned. Gresham (2009) brings attention to there being no current evidence to suggest individual components of an intervention actually correlate to each other. Therefore, a commonly accepted method for determining reliability of an instrument (i.e. Cronbach's Alpha) may not provide an accurate indicator.

Factors of Implementation – Self Report Measure

The Factors of Implementation – Self Report Measure was created by the researcher in order to measure several factors of implementation hypothesized to be related to the level of treatment integrity as discussed in the literature (see Appendix H). It consisted of four sections designed to measure the Complexity factor, including complexity of the intervention, time required for implementation, number of intervention agents, and the materials and resources used for the intervention. The 17 items used to assess complexity were taken directly from Hunter's (1982) direct instruction model. Each element in Hunter's (1982) model was assessed individually on a six-point scale ranging from "Not at all Implemented" to "Implemented Completely". In addition to the

four sections, a short demographic section was provided requesting teacher name, sex, grade level taught, and number of years taught. Teachers completed the Factors of Implementation Checklist at the completion of the seven-week intervention. A reliability coefficient for the FOI checklist was established using Cronbach's Alpha and found to be at an acceptable level (.925).

Factors of Implementation Checklist – Observations

The Factors of Implementation – Observations was a shortened form of the Factors of Implementation Checklist – Self Report Measure (see Appendix I). It was designed to obtain an objective measure of the complexity of the intervention. Time required for implementation, number of intervention agents, and the materials and resources used for the intervention are not elements that can be easily observed and therefore were not included in the checklist. The Factors of Implementation – Observations was completed by the researcher once for each of the two interventions used in the study (Voyager Passport and Read Well) for comparison with teacher perceptions.

Dependent Variable

Importance of Reading

Since the transformation from the traditional discrepancy model to an RTI framework, the focus has been on academics, mainly early literacy skills and reading. The reasoning behind this focus had to do with the importance of mastering reading skills (National Institute of Child Health and Human Development, NIH, 2000). The National Reading Panel (NRP) issued a report in April 2000 after reviewing over 100,000 studies on reading and reading development, and concluded that there are five critical skills of

reading: phonemic awareness, phonics, fluency, comprehension, and vocabulary (NIH, 2000). The NRP findings had a huge impact on the development of No Child Left Behind legislation and the Reading First initiative. The Reading First initiative built upon the findings of the NRP to foster the development of early literacy and reading skills through the use of scientifically based programs in early education (NIH, 2000). Federal funds continue to focus on the prevention and early intervention of reading problems, which clearly indicates the enormity of the problem.

Early literacy skills and reading are taught to children, beginning in kindergarten and continuing through the third grade. Beginning in third grade, the focus on learning to read decreases and students are expected to read independently to learn. Therefore, students who are poor readers will not only do poorly in reading but will do poorly in other academic areas as well. Children struggling with academics often develop behavior problems and become more challenging to teach effectively (Biglan, Brennan, Foster, & Holder, 2004). An interaction occurs between academic and behavior problems and a cycle begins. Children with behavior problems fail to access the curriculum and fall behind in their academics and children who are behind in academics become frustrated by demands they cannot meet, causing more behavior problems (Biglan et al., 2004). Taking it one step further, students who get caught up in the cycle and do not receive proper intervention are likely to end up needing special education services due to observed skill deficits. The Oregon Social Learning Center has actually identified six key risk factors that are strongly associated with becoming a juvenile offender, and number five on this list is having received special education services at one time or another (Sprague & Walker, 2005). Clearly, prevention of and early intervention in

academic (especially reading) problems is essential to prevent countless problems in the future, and this led to the researcher's decision to focus on treatment integrity of reading interventions.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS)

As described above, the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) (Dynamic Measurement Group, Inc., 2009) are assessment tools that are low-cost, short, quick, and direct ways of measuring skill accuracy and fluency. DIBELS (Dynamic Measurement Group, Inc., 2009) data, more specifically Oral Reading Fluency (DORF) data, were collected for all students. Reliability estimates remain adequate for the DORF range from 0.89 to 0.98 (Baker, Smolkowski, Katz, Fien, Seely, Kame'enui, et al., 2008; Roberts, Good, & Corcoran, 2005) according to recent research. Concurrent (0.82) and predictive validity (0.63, 0.71, 0.72) with the Stanford Achievement Test, 10th edition were also reported by Baker et al., 2008). Schilling, Carlisle, Scott, and Zeng (2007) reported concurrent validity estimates (0.61, 0.69, 0.71, 0.74, 0.75) and predictive validity estimates (0.61, 0.61, 0.63, 0.69, 0.69) with the Iowa Tests of Basic Skills. DORF data were also collected at the end of the seven-week intervention period by the intervention agents on students whose performance fell below the established criterion level at the September benchmark assessment.

A criticism of DIBELS (Dynamic Measurement Group, Inc., 2009) and CBM probes in general, is that they are grade-dependent, making it difficult to compare gains between grade levels. For example, a second-grade student will make greater gains on an oral reading fluency probe from week 1 to week 2 than a third-, fourth-, or fifth-grade student would make in the same period using their respective oral reading fluency grade

level probes. To remedy this, the researcher used the documented Rate of Improvement (ROI) for each independent grade level at their respective percentiles, according to pre-intervention scores from the September benchmarking period, to develop goal scores for each participant. The participant's post-intervention score was then compared to the created goal score to determine whether the goal was met. For example, participant A, a second-grade student, received a score of 15 on oral reading fluency pre-intervention and a score of 22 on oral reading fluency post-intervention (seven weeks later). The expected rate of improvement for a second-grade student scoring at the 10th percentile is 0.8. The ROI of 0.8 was multiplied by the seven weeks of intervention, which equals 5.6. Therefore, the goal score for participant A is 20.6 (15 + 5.6) words read per minute on an oral reading fluency second-grade probe. Participant A's post score of 22 exceeded the goal score of 20.6 by 1.4. The 1.4 was then used as the standardized variable to compare overall gains made between grade levels.

CHAPTER 4

RESULTS

Purpose of the Study

The purpose of the current study was to examine the relationship between factors of implementation, treatment integrity, and reading fluency. Specifically, the study examined the relationship between complexity and acceptability factors, treatment integrity of Tier II reading interventions, and student scores on DIBELS oral reading fluency. After initial data screening, the following hypotheses were tested. The first research question explored the direct relationships between factors of intervention implementation, treatment integrity and reading fluency. It was hypothesized that low scores for the complexity factor would predict a high level of treatment integrity and that high scores for the acceptability factor would predict a high level of treatment integrity. High scores for treatment integrity would predict greater gains in reading. The second research question explored how treatment integrity mediated these relationships. It was hypothesized that low scores for the complexity factor would predict greater gains in reading as mediated by treatment integrity and that high scores for the acceptability factor would predict greater gains in reading as mediated by treatment integrity. To test these hypotheses, a technique that involved the estimation of presumed causal relationships among the observed variables, known as a path analysis (Kline, 2005), was conducted using EQS 6, a structural equation program. Path analysis is an extension of regression modeling used to test the fit of the correlation matrix against two or more causal models compared by the researcher (Kline, 2005). Four variables (complexity, acceptability, integrity, and outcomes) specified the path model based upon hypotheses of the expected

causal relationships. Although variables were placed into a path model based on expected causal relationships, the analysis does not explain causality between variables but rather correlations between variables.

Descriptive Statistics and Data Screening

Prior to conducting a path analysis, several assumptions must be satisfied. First, linearity is assumed, which implies that relationships among variables are linear. Second, interval level data is assumed, which means data should be at interval scale. Third, uncorrelated residual term is assumed, which means the error term should not be correlated to any variable. Fourth, disturbance terms should not be correlated to endogenous variables is assumed. Fifth, low multicollinearity is assumed. A sixth assumption is the path analysis should not be under-identified; exactly identified or over-identified models are good for path analysis. Data analysis in EQS 6 confirmed the assumptions were satisfied. Seventh, adequate sample size is needed to assess significance. Kline (1998) recommends at least 10 times as many cases or parameters (ideally 20 times) as sufficient. Also, according to an a priori power analysis where power was set at 0.80 and alpha at 0.05, the minimum number of participants needed for a medium effect size using four factors was 84. Therefore, the 122 participants used in the sample satisfy the minimum requirements of the assumption. A large effect size is typically desired but, owing to the exploratory nature of the present study, a medium effect size is more appropriate. Lastly, the same sample size is required for all regressions. This assumption was satisfied by removing variables with missing values from the data set. A total of 38 cases were removed owing to missing variables.

Table 1 contains data on the means, skewness, kurtosis, and standard deviations of each of the four variables: complexity, acceptability, integrity, and outcomes. The complexity variable contained four subscales: complexity, time, number of intervention agents, and materials and resources. Table 2 contains data on the means and standard deviations of the four subscales of the complexity variable. Table 3 contains data on the ranges of each scale and subscale. Table 4 contains bivariate correlations of the complexity subscales correlated with each other, correlated with treatment integrity and correlated with student reading outcomes.

Table 1: Univariate Statistics

Variable	*n*	Mean	Skewness	Kurtosis	Standard Deviation
Complexity	122	413.6475	1.3880	2.1080	99.6468
Acceptability	122	64.0820	-.6803	-.3984	8.8734
Integrity	122	55.5492	.7584	.0938	4.9578
Outcomes	122	5.0164	.1721	.0493	11.6535

Table 2: Complexity Subscales – Means and Standard Deviations

Subscale	*n*	Mean	Standard Deviation
Complexity	122	83.47	9.62
Time	122	327.97	94.83
Agents	122	1.48	.911

Table 3: Ranges: Scales and Subscales

Variable	*n*	Range	Minimum	Maximum
Complexity	122	434	259	693
Complexity	122	36	63	99
Time	122	410	190	600
Agents	122	4	1	5
Acceptability	122	34	43	77
Integrity	122	19	48	67
Outcomes	122	60	-28.6	32.10

Table 4: Bivariate Correlations

	Complexity	Time	Agents	Materials	Integrity	Outcomes
Complexity						
Pearson Correlation	1	.479**	.230*	-325**	-.222*	-.063
Sig. (2-tailed)		.000	.011	.000	.014	.492
Time						
Pearson Correlation	.479**	1	-.120	-.233**	-.221*	-.019
Sig. (2-tailed)	.000		.189	.010	.014	.833
Agents						
Pearson Correlation	.230*	-.120	1	.454**	-.010	-.265**
Sig. (2-tailed)	.011	.189		.000	.914	.003
Materials						
Pearson Correlation	-.325**	-.233**	.454**	1	.242**	-.203*
Sig. (2-tailed)	.000	.010	.000		.007	.025
Integrity						
Pearson Correlation	-.222*	-.221*	-.010	.242**	1	.228*
Sig. (2-tailed)	.014	.014	.914	.007		.011
Outcomes						
Pearson Correlation	-.063	-.019	-.265**	-.203*	.228*	1
Sig. (2-tailed)	.492	.883	.003	.025	.011	

Inter-observer Agreement

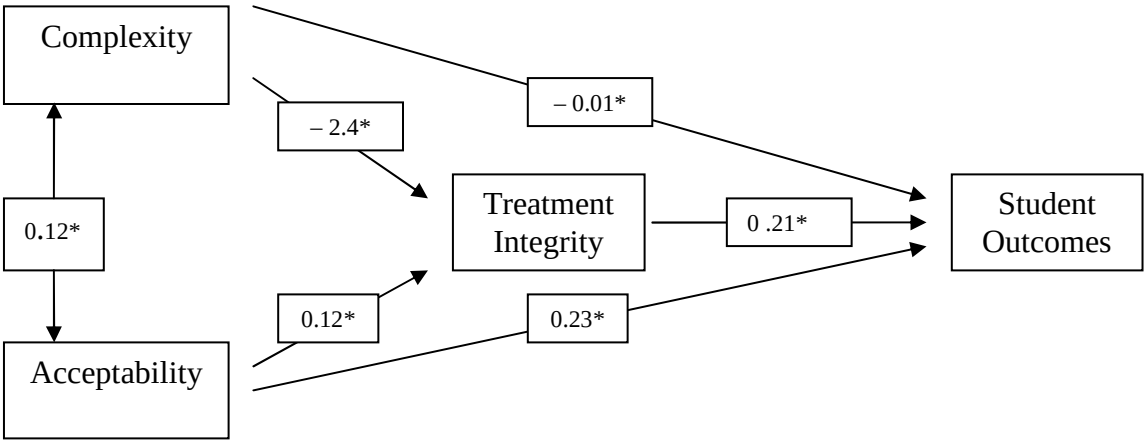
It was proposed that inter-observer agreement be assessed on 20% of the days that observation data were collected. Owing to schedule changes, however, second observers were available for only 12% of the proposed observations. A point-by-point agreement ratio was calculated to assess inter-observer agreement of scores on the Treatment Integrity Observation tool. Agreement ratios were expected to be low, owing to the 6-point scale used to measure intervention integrity. Agreement ranged from 62% to 92%, with a mean of 74%. The following formula was used to calculate percentage of agreement: number of agreements/number of agreements + number of disagreements x 100.

Interpretation: Path Analysis Model

Results from the path analysis indicate both significant direct and indirect effects within the model are consistent with both hypotheses proposed in the present study: $c^2 = (6, N = 122) = 0.000$, $P < 0.05$. Direct and indirect effects of the variables are shown in Figure 1. The proposed hypothesis for question one was supported by the model through significant direct effects. An inverse relationship was found between the complexity factor and treatment integrity. Teachers that rated intervention complexity to be low implemented the intervention with greater integrity ($r = – 0.24$, $P < 0.05$). A direct relationship was found between the acceptability factor and treatment integrity. Teachers that rated acceptability of the intervention to be high implemented the intervention with greater treatment integrity ($r = 0.12$, $P < 0.05$). Six percent of the variance in treatment integrity was accounted for by complexity and acceptability. A direct relationship was also found between treatment integrity and reading fluency. Teachers that implemented

the intervention with greater integrity displayed greater gains in reading ($r = 0.21$, $P < 0.05$). The proposed hypothesis for question two was supported by the model through significant indirect effects. An inverse relationship was found between the complexity factor and reading fluency. Teachers that rated intervention complexity to be low displayed greater gains in reading as mediated by treatment integrity ($r = – 0.01$, $P < 0.05$). A direct relationship was found between the acceptability factor and reading fluency. Teachers that rated acceptability of the intervention to be high displayed greater gains in reading fluency as mediated by treatment integrity ($r = 0.23$, $P < 0.05$). Ten percent of the variance in student outcomes was accounted for by complexity, acceptability, and treatment integrity.

Figure 2: Path Analysis Model



* = p < .05

CHAPTER 5

DISCUSSION

Introduction to the Problem

Research in the area of treatment integrity and factors that influence treatment integrity still remains limited. However, it is expected that recent changes in legislation and the adoption of an RTI framework will increase the focus on research and practice of treatment integrity. Promising indications of a greater focus have already been noted with the most recent publication of the National Association of School Psychologists dedicating a special series to treatment integrity. The purpose of the special series was to review and synthesize current research and to address issues related to treatment integrity in both measurement and the use of treatment integrity data for practice. The recent focus on treatment integrity was best noted by Hagermoser Sanetti and Kratochwill (2009) citing changes in federal legislation (i.e. NCLB, 2002; IDEA, 2004), professional organizations' position statements (i.e. National Association of School Psychologists, 2005), added inclusion of treatment integrity data needed for research proposals (i.e. Institute for Education Sciences, 2009), and movements toward RTI in practice. Previous research has indicated several factors of implementation to be related to treatment integrity. The purpose of this investigation was to explore these factors (complexity and acceptability) and their relationship to treatment integrity of intervention implementation and ultimately student outcomes (i.e. reading fluency).

Despite several limitations, valuable conclusions can be drawn from this study that will contribute to the present and future research conducted in the area of treatment integrity. Findings from this study further support the existing literature on factors

related to treatment integrity, as well as, provide important additions to the literature. This chapter provides an overview of the study, interpretation of the analysis, limitations of the study, implications for school psychologists, and future research.

Overview of the Study

The purpose of this study was to examine the relationship between factors of implementation, treatment integrity, and student's reading fluency. Evaluation of these elements was conducted at an elementary school in the southeastern part of the United States and included participants in grades 2 through 5. General education teachers and instructional aides answered pre- and post-intervention questionnaires, allowed observations of intervention groups, and provided pre- and post-intervention DORF data for student participants. A total of 122 students who met the criteria were included as participants in the study. A path analysis model was constructed based on hypotheses of the relationships between factors of implementation (complexity and acceptability), treatment integrity, and student outcomes (DORF).

Interpretation of the Analysis

Question 1: Direct Effects

The first research question in the study was: What are the direct relationships between implementation factors, treatment integrity and reading fluency? To address this question, data were collected for each of these elements. To assess factors of implementation, teachers and instructional aides filled out questionnaires both pre (i.e. AIM) and post–intervention (i.e. AIM & FOI-Self) using instruments adapted from standardized instruments. To assess treatment integrity, school psychologist interns and the researcher conducted structured observations using an instrument adapted from a

standardized instrument. To assess reading fluency, data were collected pre- and post-intervention using a published, standardized instrument targeting reading fluency skills (i.e. DORF). Multiple sources of data (i.e. self report, observations, and permanent product) were collected based upon the recommendation by several researchers (i.e. Gansle & Noell, 2007; Gresham, 1996) owing to disadvantages of both direct and indirect forms of assessment. Data were analyzed using a path analysis structural equation model.

The results demonstrated direct relationships between complexity and treatment integrity (inverse relationship), complexity and reading fluency (inverse relationship), acceptability and treatment integrity, acceptability and reading fluency, and treatment integrity and reading fluency. All relationships were determined to be significant at the 0.05 level. Consistent with previous research, the more complex the intervention was rated by teachers, the lower the level of treatment integrity. A study conducted by Becker and Carnine (1980) also found a complex Direct Instruction model to be related to lower levels of treatment integrity. However, the complexity factor developed for this study included several factors previously found to be related to treatment integrity: complexity, time, resources, and number of agents. It was hypothesized that these factors were all interrelated to each other and therefore do not need to be measured separately. Even when several factors were combined into the one factor (complexity), findings of the present study indicate the more time needed for implementation, the more unfamiliar the resources, and the more people involved in implementation, all continue to support previous literature in that they are related to lower levels of treatment integrity. Previous research has proposed an interaction between complexity and time, as interventions that

are more complex require more time for implementation. Although this link has been proposed, to date it has never been supported by research. The present study provides support to the notion that complexity, time, resources, and number of treatment agents are possibly linked. The present findings about the relationship between time and treatment integrity are consistent with Happe's (1982) finding of 87% of implementation agents reporting lack of time as a factor for not implementing a plan with integrity. Additionally, resources indicated by teachers as beyond the scope of their use within the general education classroom were related to lower levels of treatment integrity as previously found in review of the literature (Gresham, 1989). Lastly, the more intervention agents involved with implementation (i.e. prepared materials, provided implementation, collected progress monitoring data) were related to lower levels of treatment integrity as indicated previously by Gresham (1989).

In addition to providing findings consistent with past research on the relationship between complexity and treatment integrity, the present study contributed to new research by providing a structured way for measuring treatment complexity through the use of Hunter's (1982) guidelines for direct instruction. Previously, research has not defined interventions as complex through the use of a structured measurement tool, but rather as interventions consisting of multiple elements. The Factors of Implementation checklist developed for the present study allowed for direct measurement of specific elements within an intervention including: identifying objectives, knowledge of standards, anticipatory set or hook, teaching, which includes input, modeling, and checking for understanding, guided practice, closure, and independent practice (Hunter, 1982). The present study found through the use of the Factors of Implementation

checklist that the more elements included in the intervention as rated by the intervention agent, the higher the level of complexity within the intervention. Although the findings in the present study are promising, additional research is necessary to further explore the use of this measurement tool for determining the complexity of an intervention. The more complex an intervention, the lower the level of treatment integrity for intervention implementation has consistently been shown in the research and was again supported by the present study. However, it is still unknown the exact level of complexity that is related to treatment integrity. The Factors of Implementation checklist or another structured tool for the measurement of treatment integrity may be useful for determining this cutoff level.

The relationship between complexity and student outcomes is a new addition to treatment integrity research by the present study. Research has previously linked complexity to treatment integrity and treatment integrity to student outcomes, but this was the first time a direct relationship has been found between complexity and student outcomes (i.e. reading fluency). When complexity was rated low, reading fluency was high even with treatment integrity removed from the model. A possible benefit to the understanding of this direct relationship may contribute to a greater focus on manipulating complexity of an intervention, rather than the current focus on measurement of treatment integrity. Measurement of treatment integrity may not be necessary if the findings of the present study are replicated in future research.

Also consistent with past research, the higher acceptability of the intervention was rated, the higher the level of treatment integrity. Similar to the complexity factor, another addition to previous research was the combination of several factors related to treatment

integrity to develop the acceptability factor. The acceptability factor for the present study included treatment acceptability, severity of the problem, and perceived effectiveness of the intervention. Findings support previous research and indicated the more acceptable the intervention, the more severe the problem, and the greater the perceived effectiveness of the intervention, the higher the level of treatment integrity. Intervention acceptability research indicated the more acceptable the agent rated the intervention the more likely it was implemented with integrity. Braukmann et al. (1976) and Wolf (1978) found interventions that were socially validated or socially acceptable by intervention agents were more likely to be implemented with integrity, as was found in the present study. Previous research also indicated that the more severe a problem, the higher the level of treatment integrity (Elliott et al., 1984), which were consistent with the findings of the present study. In the study conducted by Elliott et al. (1984) intervention agents rated the most complex interventions as the most acceptable treatment for the most severe behavior. Additionally, intervention agents rated the most time intensive behaviors as the most acceptable treatment for the most severe behavior. These results indicate that severity of the problem as a factor overrides complexity or time as factors when rating treatment acceptability. Also, consistent with previous research (i.e. Gresham, 1989; Shinn et al., 1997) intervention agents that perceived the intervention to be effective were related to higher levels of treatment integrity. Shinn et al. (1997) demonstrated the use of performance feedback to provide support for the relationship between the factor of perceived effectiveness and reading fluency, showing increased reading fluency with performance feedback of intervention effectiveness. The Shinn et al. (1997) study provided support to the findings of the present study which indicated a relationship

between treatment acceptability and reading fluency. Findings of the present study indicated when acceptability was rated high; reading fluency scores were high even with treatment integrity removed from the model. Although the current study did not measure the relationship between student performance feedback and acceptability formally, performance feedback was available to teachers who collected DORF data either every week or every two weeks and may have been related to teachers' overall level of acceptability.

Current findings indicated that the relationship between treatment integrity and reading fluency was also found to be consistent with previous research (Gresham, 1996). When treatment integrity was high student outcomes were also high, specifically manifested as greater gains in reading fluency. Teachers were more likely to implement interventions with integrity when they perceived the intervention to be less complex and more acceptable. Interventions that were implemented with greater integrity yielded greater gains in reading fluency. Owing to complexity and acceptability as being self-report measures, the important factor for determining the level of treatment integrity may be teacher perceptions of these factors. The relationships between factors of implementation and treatment integrity continue to be established in the literature, including the present study. These relationships should be further examined in the literature. Additionally, the measurement tools developed in the present study provided a systematic, structured way to measure treatment complexity, treatment acceptability, and treatment integrity. Previous research suggested development of such measurement tools to provide a way for researchers to synthesize treatment integrity research across all types

of interventions (e.g. academic and behavior) and the present study provided promising measurement tools that can be used for a variety of types of interventions.

Question 2: Indirect Effects

The second research question was: How does treatment integrity mediate these relationships? To address this question, data were collected for each of these elements as they were described in the previous section. To assess factors of implementation, teachers and instructional aides filled out questionnaires both pre- and post-intervention using instruments adapted from standardized instruments. To assess treatment integrity, school psychologist interns and the researcher conducted structured observations using an instrument adapted from a standardized instrument. To assess reading fluency, data were collected pre- and post-intervention using a published, standardized instrument targeting reading fluency skills. Data were analyzed within the same path structural equation model.

Results indicated significant indirect relationships among the variables which were new findings to contribute to the existing literature base of treatment integrity. Although previous research did not focus on the relationship between factors of implementation, treatment integrity, and reading fluency as a possible model, each of these relationships was supported by previous research independently. As discussed previously in this section, complexity and acceptability factors have been found to be related to treatment integrity and treatment integrity has been found to be related to student outcomes (e.g. reading fluency, behavior) (e.g. Becker and Carnine, 1980; Greaham, 1989; Elliott et al., 1991). Identifying a model that shows a relationship between all of these factors provides new information to the existing literature to further

the understanding of treatment integrity. In the present study, complexity was found to be inversely related to reading fluency mediated by treatment integrity. Acceptability was also found to be related to reading fluency mediated by treatment integrity. Results were consistent with hypotheses and the basis for the construction of the path model. When teachers perceived the intervention to be less complex, treatment integrity was high and greater gains in reading fluency were achieved. When teachers rated the intervention as more acceptable, treatment integrity was high and greater gains in reading were achieved. The current study added to existing literature by presenting a model to better understand the relationship among factors of implementation, treatment integrity, and reading fluency. The model proposes that complexity of interventions and acceptability influence treatment integrity and the level of treatment integrity influences reading fluency.

Limitations

Although the study contributes to the literature in the area of treatment integrity, the results must be interpreted with caution due to limitations in the study. There are several limitations, including threats to validity as outlined by Cook and Campbell (1979), to address that may have influenced the findings of the current study. Limitations related to threats to internal validity include recruitment and attrition of participants. One teacher was unwilling to participate in the study because it was the first time that she had implemented the intervention and did not feel comfortable having someone observe. All the first-grade teachers decided not to participate, citing lack of time to fill out questionnaires. Five teachers did not complete the study: two had behavior problems with the students in their intervention groups and three did not

complete the post-intervention questionnaires. Losing these intervention groups, especially for reasons possibly affecting implementation issues, may have had an influence on the results of the study. Selection was already a possible threat to the study's internal validity owing to participants being drawn from a single elementary school, making it difficult to generalize the results to other populations. Teachers refusing to participate in the study, based on two factors previously noted in the research as being related to treatment integrity (i.e. unfamiliarity of resources/materials and time) indicate a clear selection bias. It is likely that these factors would have influenced the level of treatment integrity of interventions had the teachers taken part in the study. This may have also significantly affected the results. Additionally, teachers that dropped out of the study due to behavior issues may have also demonstrated lower levels of treatment integrity. Although behavior issues were not mentioned in the literature as related to treatment integrity, it is possible that if students are misbehaving, it may not be possible to implement the intervention completely or with integrity.

External validity is also threatened owing to interaction of treatment and selection and interaction of treatment and setting. The selection bias noted above and the single setting in which the study took place make it difficult to generalize the findings to other populations and settings, as the population of the participants of the study and the setting in which the study took place were not a representative sample of overall populations or settings. It is difficult to conclude whether the findings from this study would have been found for different populations or settings. Although this is a clear limitation of the study, replication with other populations and settings is possible and recommended for future research, which may support the findings from the current study.

The school followed a standard protocol model of service delivery, which meant that each student displaying a deficit in reading was automatically put into a reading group with a prescribed intervention. Although the standard protocol approach is supported in the research as a viable option for the delivery of services under an RTI framework, this is a limitation for the purposes of this study because only two interventions were used, thus limiting the variability. Generalizability of these results to other participants and settings is severely limited owing to the use of reading interventions only and the drawing of participants from the same elementary school.

Construct validity is also threatened owing to hypothesis guessing and confounding constructs. A limitation for collecting self-report data or conducting direct observations is the participants may answer questions or act in ways based on their impression of what the researcher expects. It is possible their answers on the survey were not accurate, but rather based upon their understanding of what the researcher expects. Teacher behaviors in the classroom during direct observations may also have been different than when the researcher or data collectors were not in the classroom owing to reactivity. It is also possible that teachers wanted to present a positive impression, by more closely adhering to the intervention protocol when the researcher and data collectors were present.

There are several possible confounding constructs that may have had an influence on the results. The participating elementary school was a Needs Improvement – Year 2 School receiving a Reading First grant. It was only the second year that the school had had problem-solving teams in place and benchmarking in reading for all grades. It was also the first year that the school had used the results of the benchmark assessments to

inform instruction and interventions. Additionally, a new reading program, a new math program, and new pre-referral interventions were implemented at the beginning of the school year. Considering the many new initiatives and procedures, and that relatively inexperienced staff were navigating within a response-to-intervention framework, it is possible that implementation integrity may have also been influenced in this way.

Other limitations are related to measurement issues within the study. As mentioned in the methods section, a nine-week intervention period was originally anticipated but, owing to the 12-month school schedule, adjustments in research design needed to be made. The entire school operates on a "track" schedule, which means that the students were out of school for a five-week period beginning at week eight of the intervention period. It was therefore decided to shorten the intervention period to seven weeks instead of having a five-week break with no intervention during those five weeks. Shortening the intervention period certainly lessened the number of days the intervention was implemented and the time available to allow growth in reading.

Implications for School Psychologists

Past research shows continued support for several factors that affect the level of treatment integrity for implementation of interventions. The present study not only found additional support for these same factors affecting the level of treatment integrity, but examined these factors in a new way by combining the individual factors into two groups: complexity and acceptability. Combining these factors has offered a simplified alternative, which may be beneficial for future research. It may not be necessary to spend the time and resources analyzing each individual factor in depth but rather focus on the type of factor, which provides the same information more efficiently. The complexity

factor created for the present study includes individual variables directly related to the implementation of an intervention that are concrete and easily manipulated, such as complexity, time, materials and resources, and additional agents involved. The acceptability factor created for the present study includes perceived effectiveness of the intervention, acceptability by the teacher, and severity of the problem. These are the internal feelings that teachers have, which are more abstract and not as easily manipulated. However, teachers were the primary reporters for both factors, implicating a possible relationship between the two factors. It is possible that the factors included in complexity are not related to the level of treatment integrity but rather to teacher perception of these complexity variables. Certainly this should be a focus of future research.

The results of the present study, along with past research to support these findings, suggest a continued focus on these factors for future research, but also for designing interventions and intervention plans. The relationship between the factors of implementation (complexity and acceptability) and treatment integrity has continually been supported through research, including the present study. Perhaps it is time for research to change focus from identification of these factors to the manipulation of these factors to determine cutoff points for which these factors are related to treatment integrity. Regardless of the theoretical implications mentioned above, the significant role that teacher perceptions play in implementing interventions has been demonstrated. Therefore, working closely with teachers to ensure positive perceptions and feelings about factors relating to intervention implementation will be beneficial for students. Teachers who accept the intervention and believe it to be effective will be more likely to

have higher levels of treatment integrity, which in turn will increase positive student outcomes. It is possible that teachers who participate in the design and selection of the intervention will have more positive perceptions about the intervention and therefore increase the likelihood of higher treatment integrity. These are the types of hypotheses that should drive future research and ultimately the development of future interventions.

Directions for Future Research

Educational research should continue to explore other possible factors affecting level of treatment integrity of interventions, especially since the research continues to show a direct relationship between treatment integrity and reading fluency. Research of treatment integrity would benefit from replication of the present study with different populations, grade levels, and types of intervention so that results may be generalized. The present study chose to focus on a small number of the factors presented by previous research, but other factors affecting treatment integrity need to be examined as well. Additional factors that could be considered include teaching style/attitude, relationship between the teacher and the students, size of intervention group, and possible behavior issues. Perhaps the other teacher variables mentioned could be included in the acceptability factor created in the present study.

Currently, there are few measurement tools used to assess treatment integrity and factors found to be related to treatment integrity, and even fewer that are technically adequate. The few tools mentioned in the literature are specific and measure different facets of treatment integrity. Development of a tool that could be used across intervention types would be a valuable asset and would allow meta-analysis of smaller studies focusing on treatment integrity. Treatment integrity research would benefit from

the development of a valid and reliable generic measurement tool to provide consistency for treatment integrity research and possibly aid in the establishment of validated cutoffs for the amount and quality of instruction (Schulte et al., 2009). Also, Schulte et al. (2009) propose that a generic tool could help to determine the number and length of observation sessions necessary to depict a representative sample and provide an accurate estimate of treatment integrity. In practical settings, treatment integrity is less likely to be a priority owing to an inability to control extraneous variables, and treatment integrity will most likely be low (Schulte et al., 2009). Therefore, not only is it necessary to develop psychometrically sound assessment tools, but tools that are efficient and feasible in applied settings (Hagermoser Sanetti & Kratochwill, 2009) should be at the forefront.

Summary

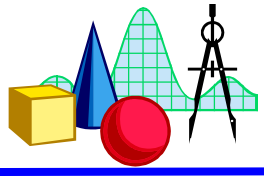
In summary, this research adds to the current literature of treatment integrity and the factors related to intervention implementation. This study demonstrated direct relationships between treatment complexity and acceptability with treatment integrity, treatment integrity with reading fluency, and treatment complexity and acceptability with reading fluency. This study also demonstrated indirect relationships between treatment complexity and acceptability with reading fluency mediated by treatment integrity. Results suggest that when intervention agents perceive an intervention to be more complex, they are less likely to implement the intervention with integrity. Findings also suggest that when an intervention is perceived as acceptable by an intervention agent, they are more likely to implement it with integrity. Furthermore, when interventions are implemented with integrity, students achieved greater gains in reading. The importance of continued research in the area of treatment integrity is immense. It is hoped that the

current legislation and a continued focus on response to intervention continue to be the driving force leading the educational community to develop the necessary tools and knowledge for implementing interventions with integrity, ultimately promoting more positive outcomes for students.

APPENDIX A

DISTRICT INTERVENTION PROTOCOL

Response to Intervention
Intervention Assessment Protocol
Research-based Standards



Student	**Student ID Number**	**Recorder**
Intervention Target Concern		
Performance Indicator (PI) *(Skill-specific element of the Target Concern to be observed and measured)*		
Intervention Description		

Intervention Planning & Development

Y	**N**	**Date**	**There is evidence that…**
			1. Analysis of the *Target Concern* was based upon a *functional assessment* of the problem.
			2. A representative skill of the *Target Concern* was identified as the *Performance Indicator.*
			3. A baseline level of functioning was determined for the *Performance Indicator.*
			4. Initial intervention instruction matched the student's baseline level of functioning.
			5. Timelines were appropriate for determining *skill-specific* performance changes.
			6. An appropriate *Target Goal* was established given the baseline level of functioning.
			7. A measurement and data recording plan was developed.

Integrity of Intervention Implementation

Y	**N**	**Date**	**There is evidence that…**
			1. The intervention plan was documented in writing. *(If "Yes," attach a copy to this document)*
			2. Responsibilities were explained to all participants, including the student.
			3. Materials and resources were obtained and appropriately utilized.
			4. Performance data were appropriately collected.
			5. Performance data were charted or graphed.
			6. The intervention was implemented as designed and planned.

Intervention Acceptability

Y	N	
		The preponderance of evidence indicates that the intervention was planned, developed and implemented according to research-based standards.

Explanation of Deficiencies (If "Yes," is indicated, then briefly explain any item above that was checked "No.")

APPENDIX B

DISTRICT INTERVENTION PLAN

INTERVENTION PLAN

STUDENT NAME:		**DATE:**	
STUDENT NUMBER:		**BIRTHDATE:**	
TEACHER/TRACK:		**AGE:**	
SIP CASE MANAGER:		**ELP:**	

Intervention Plan # (circle one)	**1**	**2**	**3**	**4**	**5**
To–From Dates					

BASELINE/TARGET CONCERN(S): (Describe in specific, observable & measurable terms)

_____

_____

_____

INTERVENTION GOAL(S): (Must be specific, observable & measurable)

_____

_____

_____

_____

INTERVENTION PLAN SUMMARY: (Identify specific interventions that will be used to increase or decrease the target concern)

_____

_____

_____

_____

INTERVENTION TIME REQUIRED TO MEET GOAL:

Number of instructional sessions per week:	Number of minutes per instructional session:

OTHER FACTORS: (Identify any necessary resources, materials, setting, sequencing of intervention steps, parent participation and person(s) responsible)

_____

_____

HOW WILL PROGRESS BE MEASURED? (Attach specific assessment if possible)

_____

_____

OUTCOME: TC=Target Concern: Date of Decision:_____

___TC resolved (SIP terminated) ___TC being resolved (SIP continued) _____TC unresolved (explain)-

_____

APPENDIX C

DISTRICT INTERVENTION LOG

INTERVENTION LOG

STUDENT NAME:	**DATE:**
STUDENT NUMBER:	**BIRTHDATE:**
TEACHER/ROOM:	**STUDENT AGE:**
CASE MANAGER:	**SCHOOL:**
ELP:	**GRADE/TRACK:**

Date	Minutes	Teacher: Student Ratio	Instruction/Intervention	Comments/Progress	Initials

APPENDIX D

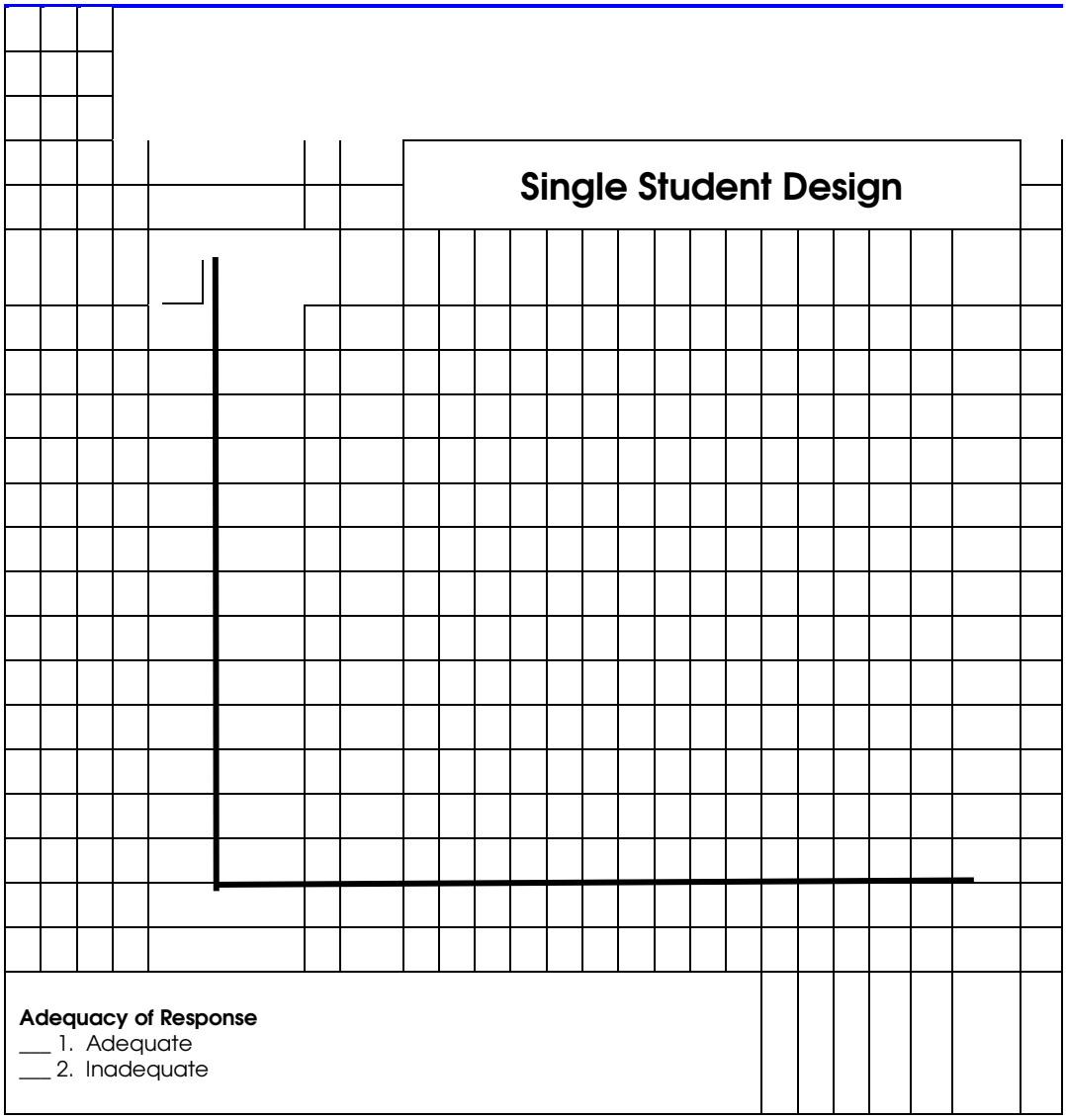
DISTRICT INTERVENTION GRAPH

Intervention Graph

Student Name: _____ Date: _____

Performance Indicator (Y-axis): _____

Time Interval (X-axis): _____



APPENDIX E

ACCEPTABILITY OF INTERVENTION (AIM) – PRE-INTERVENTION SURVEY

1. How would you rate the severity of the student's target problem?

Not at all severe	**1**	**2**	**3**	**4**	**5**	**6**	Very severe

Regarding the intervention used, please circle the number that best describes your agreement/disagreement with each statement, using the following scale:

1 = strongly disagree	2 = disagree	3 = slightly disagree	4 = slightly agree	5 = agree	6 = strongly agree

2.	I like the procedures in this intervention.	**1**	**2**	**3**	**4**	**5**	**6**
3.	This intervention is a good way to handle the problem.	**1**	**2**	**3**	**4**	**5**	**6**
4.	Overall, this intervention will be beneficial for the child.	**1**	**2**	**3**	**4**	**5**	**6**
5.	I have the time needed to implement this intervention.	**1**	**2**	**3**	**4**	**5**	**6**
6	I have the materials needed to implement this intervention.	**1**	**2**	**3**	**4**	**5**	**6**
7.	I have the support needed to implement this intervention.	**1**	**2**	**3**	**4**	**5**	**6**
8.	I would suggest the use of this intervention to other teachers.	**1**	**2**	**3**	**4**	**5**	**6**
9.	The child's problem is severe enough to warrant the use of this intervention.	**1**	**2**	**3**	**4**	**5**	**6**
10.	The intervention is a fair way to handle the child's problem.	**1**	**2**	**3**	**4**	**5**	**6**
11.	The intervention is appropriate for a variety of children.	**1**	**2**	**3**	**4**	**5**	**6**

12. How would you rate the effectiveness of the intervention?

Not at all	**1**	**2**	**3**	**4**	**5**	**6**	Extremely

13. To what extent will you able to implement the intervention as designed?

Not at all	**1**	**2**	**3**	**4**	**5**	**6**	Exactly as planned

APPENDIX F

INTERVENTION INTEGRITY CHECKLIST

Observation #: _____
Name of Observer: _____
Date of Observation: _____
Location of Observation: _____
Start Time: _____
End Time: _____

Not at all Implemented	**1**	**2**	**3**	**4**	**5**	**6**	Implemented Completely

	Intervention Component						
1.	The intervention agent introduces the intervention to the student.	1	2	3	4	5	6
2.	The intervention takes place in a suitable environment (i.e. one with minimal disruptions).	1	2	3	4	5	6
3.	The intervention agent implements each step of the intervention.	1	2	3	4	5	6
4.	The intervention agent implements the intervention for the correct length of time.	1	2	3	4	5	6
5.	The intervention agent uses the proper materials.	1	2	3	4	5	6
6.	The intervention agent provides the student with examples.	1	2	3	4	5	6
7.	The intervention agent provides the student with non-examples	1	2	3	4	5	6
8.	The intervention agent models the correct behavior for the student.	1	2	3	4	5	6
9.	The intervention agent checks for understanding before proceeding to the next step.	1	2	3	4	5	6
10.	The intervention agent provides opportunities for the student to practice the newly learned skill.	1	2	3	4	5	6
11.	The intervention agent provides support based on the student's need through scaffolding.	1	2	3	4	5	6
12.	The intervention agent assesses the student's progress.	1	2	3	4	5	6
13.	The intervention agent records the student's progress.	1	2	3	4	5	6

APPENDIX G

ACCEPTABILITY OF INTERVENTION (AIM) POST-INTERVENTION SURVEY

1. How would you rate the severity of the student's target problem?

Not at all severe	**1**	**2**	**3**	**4**	**5**	**6**	Very severe

Regarding the intervention used, please circle the number that best describes your agreement/disagreement with each statement, using the following scale:

1 = strongly disagree	2 = disagree	3 = slightly disagree	4 = slightly agree	5 = agree	6 = strongly agree

2.	I liked the procedures used in this intervention.	**1**	**2**	**3**	**4**	**5**	**6**
3.	This intervention was a good way to handle the problem.	**1**	**2**	**3**	**4**	**5**	**6**
4.	Overall, this intervention was beneficial for the child.	**1**	**2**	**3**	**4**	**5**	**6**
5.	I had the time needed to implement this intervention.	**1**	**2**	**3**	**4**	**5**	**6**
6	I had the materials needed to implement this intervention.	**1**	**2**	**3**	**4**	**5**	**6**
7.	I had the support needed to implement this intervention.	**1**	**2**	**3**	**4**	**5**	**6**
8.	I would suggest the use of this intervention to other teachers.	**1**	**2**	**3**	**4**	**5**	**6**
9.	The child's problem was severe enough to warrant the use of this intervention.	**1**	**2**	**3**	**4**	**5**	**6**
10.	The intervention is a fair way to handle the child's problem.	**1**	**2**	**3**	**4**	**5**	**6**
11.	The intervention would be appropriate for a variety of children.	**1**	**2**	**3**	**4**	**5**	**6**

12. How would you rate the effectiveness of the intervention?

Not at all	**1**	**2**	**3**	**4**	**5**	**6**	Extremely

13. To what extent were you able to implement the intervention as designed?

Not at all	**1**	**2**	**3**	**4**	**5**	**6**	Exactly as planned

APPENDIX H

FACTORS OF IMPLEMENTATION CHECKLIST – SELF-REPORT MEASURE

Demographics Section

Teacher Name: _____

Gender (circle): M F

Current grade level you teach: _____

Number of years teaching: _____

Please circle the number to indicate the level of implementation for each element of complexity that occurred during the intervention.

Not at all Implemented	**1**	**2**	**3**	**4**	**5**	**6**	Implemented Completely

Complexity of the Treatment Elements and Definitions						
Objectives:						
a. It has been explained to the student what he/she should be able to do.	1	2	3	4	5	6
b. It has been explained to the student what he/she should understand.	1	2	3	4	5	6
c. It has been explained to the student what he/she should care about as a result of the teaching.	1	2	3	4	5	6
Standards:						
a. There has been an explanation of the type of lesson to be presented.	1	2	3	4	5	6
b. There has been an explanation of the type of procedures to be followed.	1	2	3	4	5	6
c. There has been an explanation of the behavioral expectations related to the intervention.	1	2	3	4	5	6
d. There has been an explanation of what knowledge or skills are to be demonstrated.	1	2	3	4	5	6
Anticipatory Set:						
a. The student's attention has been gained (i.e. actions and statements by the teacher relate the experiences of the student to the objectives of the lesson.)	1	2	3	4	5	6
Input:						
a. The teacher provides the information needed for students to gain the knowledge or skill through	1	2	3	4	5	6

lectures, films, tapes, videos, pictures, etc.						
Modeling:						
a. Once the material has been presented, the teacher uses it to show student examples of what is expected as an end product of his/her work.	1	2	3	4	5	6
b. The critical aspects are explained through labeling, categorizing, comparing, etc.	1	2	3	4	5	6
c. The student is taken to the application level (problem-solving, comparison, summarizing, etc.)	1	2	3	4	5	6
Checking for understanding:						
a. It has been determined whether the student has "got it" before proceeding (i.e. there was either a formal or informal way to assess the skill being taught.)	1	2	3	4	5	6
Guided practice:						
a. An opportunity has been provided for the student to demonstrate a grasp of new learning by working through an activity or exercise under the teacher's direct supervision.	1	2	3	4	5	6
Closure:						
a. The teacher has reviewed and clarified the key points of the lesson.	1	2	3	4	5	6
Independent Practice:						
a. Reinforcement practice is provided on a repeating schedule so that the learning is not forgotten.	1	2	3	4	5	6
b. The reinforcement practice is provided in enough different contexts so that the skill/concept may be applied to any relevant situation, not only the context in which it was originally learned.	1	2	3	4	5	6

Please indicate how much time you spent on each of the following activities:

Time Required for Implementation	**Time (minutes)**
Preparation for the Intervention: Includes any time that you spend photocopying, gathering materials, practicing with the materials, and organizational planning for the intervention.	
Actual Intervention: The time from when you begin the intervention with the student to the time you end.	
Collection of Progress Monitoring Data: Includes organizing and printing materials, data collection, and recording the student's progress.	

Please indicate any additional intervention agents:

Number of Intervention Agents	Name	Activities
Does anyone help you to prepare for the intervention, administer the intervention, and/or collect progress monitoring data? If yes, list their name(s) and how they help.	_____ _____	_____ _____ _____ _____

Please list all materials and resources used for the intervention and indicate whether these materials are typically or not typically used in your teaching. Typically used materials are considered to be anything you use as part of the general curriculum (i.e. it is used for all students in your classroom.

Materials/Resources Used	Typically	Not typically
a)		
b)		
c)		
d)		
e)		
f)		

APPENDIX I

FACTORS OF IMPLEMENTATION CHECKLIST - OBSERVATIONS

Please circle the number to indicate the level of implementation for each element of complexity that occurred during the intervention.

Not at all Implemented	**1**	**2**	**3**	**4**	**5**	**6**	Implemented Completely

Complexity of the Treatment Elements and Definitions						
Objectives:						
a. It has been explained to the student what he/she should be able to do.	1	2	3	4	5	6
b. It has been explained to the student what he/she should understand.	1	2	3	4	5	6
c. It has been explained to the student what he/she should care about as a result of the teaching.	1	2	3	4	5	6
Standards:						
a. There has been an explanation of the type of lesson to be presented.	1	2	3	4	5	6
b. There has been an explanation of the type of procedures to be followed.	1	2	3	4	5	6
c. There has been an explanation of the behavioral expectations related to the intervention.	1	2	3	4	5	6
d. There has been an explanation of what knowledge or skills are to be demonstrated.	1	2	3	4	5	6
Anticipatory Set:						
a. The student's attention has been gained (i.e. actions and statements by the teacher relate the experiences of the student to the objectives of the lesson.)	1	2	3	4	5	6
Input:						
a. The teacher provides the information needed for students to gain the knowledge or skill through lectures, films, tapes, videos, pictures, etc.	1	2	3	4	5	6
Modeling:						
a. Once the material has been presented, the teacher uses it to show student examples of what is expected as an end product of his/her work.	1	2	3	4	5	6
b. The critical aspects are explained through labeling, categorizing, comparing, etc.	1	2	3	4	5	6
c. The student is taken to the application level (problem-solving, comparison, summarizing, etc.)	1	2	3	4	5	6
Checking for understanding:						

a. It has been determined whether the student has "got it" before proceeding (i.e. there was either a formal or informal way to assess the skill being taught.)	1	2	3	4	5	6
Guided practice:						
a. An opportunity has been provided for the student to demonstrate a grasp of new learning by working through an activity or exercise under the teacher's direct supervision.	1	2	3	4	5	6
Closure:						
a. The teacher has reviewed and clarified the key points of the lesson.	1	2	3	4	5	6
Independent Practice:						
a. Reinforcement practice is provided on a repeating schedule so that the learning is not forgotten.	1	2	3	4	5	6
b. The reinforcement practice is provided in enough different contexts so that the skill/concept may be applied to any relevant situation, not only the context in which it was originally learned.	1	2	3	4	5	6

REFERENCES

Algozzine, B., Ysseldyke, J. E., Christenson, S., & Thurlow, M. (1983). *Teachers' intervention choices for children exhibiting different behaviors in school.* Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.

Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J. R., Kame'enui, E. J., et al. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review*, 37(1), 18–37.

Becker, W., & Carnine, D. (1980). Direct instruction: An effective approach to educational intervention with disadvantaged and low performers. In B. Lahey & A. Kazdin (Eds.). *Advances in clinical child psychology* (vol. 3) (pp. 429–473). New York: Plenum Press.

Biglan, A., Brennan, P.A, Foster, S.L., and Holder, H.D. (2004) *Helping adolescents at risk: Prevention of multiple problem behaviors*. New York: The Guilford Press.

Braukmann, C. J., Kirigin, K. A., & Wolf, M. M. (1976). *Achievement Place: The researcher's perspective.* Paper presented at the meeting of the American Psychological Association, Washington, D.C., September, 1976.

Brown-Chidsey, R. & Steege, M. (2005). *Response to intervention: Principles and strategies for effective practice.* New York: The Guilford Press.

Buck, J. & Torgeson, J. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test*. Tallahassee, FL: Florida Center for Reading Research.

Cook, T., & Campbell, D. (Eds.) (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally.

Dane, A. V. & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, 18, 23–45.

Deno, S. (1989). Curriculum-based measurement and special education services: A fundamental and direct relationship. In M. Shinn, *Curriculum-based measurement: assessing special children* (pp. 1–17). New York: The Guilford Press.

Dynamic Measurement Group, Inc. (2009). DIBELS [Curriculum Based Measurement]. Retrieved from http://www.dynamicmeasurement.org/index.html

Eckert, T. L. & Hintze, J. M. (2000). Behavioral conceptions and applications of acceptability: Issues related to service delivery and research methodology. *School Psychology Quarterly*, 15, 123–148.

Elliott, S.N. & Treuting, M.V. (1991). The Behavior Intervention Rating Scale: Development and validation of a pretreatment acceptability and effectiveness measure. *Journal of School Psychology*, 29, 43–51.

Elliott, S. N., Witt, J. C., Galvin, G., & Peterson, T. (1984). Acceptability of positive and reductive interventions: Factors that influence teachers' decisions. *Journal of School Psychology*, 22, 353–360.

Elliott, S.N., Witt, J. C., Kratochwill, T. R. (1991). Selecting, implementing, and evaluating classroom interventions. In G. Stoner, M.R. Shinn, & H. M. Walker. *Interventions for achievement and behavior problems*. Maryland: National Association for School Psychologists.

Epps, S. (1985). Best practices in behavioral observation. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology* (pp. 95–111). Washington, D.C.: National Association of School Psychologists.

Francis, D.J., Shaywitz, S.E., Stuebing, K.K., Shaywitz, B.A., & Fletcher, J.M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology*, 88, 3–17.

Fuchs, L. S., Fuchs, D., & Speece, D. L. (2002). Treatment validity as a unifying construct for identifying learning disabilities. *Learning Disability Quarterly*, 25, 33–45.

Fuchs, D., Mock, D., Morgan, P.L., & Young, C.L. (2003) Responsiveness-to-Intervention: Implication for the Learning Disabilities Construct. *Learning Disabilities Research and Practice,* 18, 3, 157-171.

Gansle, K. A. & Noell, G.H. (2007). The fundamental role of intervention implementation in assessing response to intervention. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.). *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 10–24). New York: Springer.

Glover, T. A. & DiPerna, J. C. (2007). Service delivery for response to intervention: Core components and directions for future research. *School Psychology Review*, 36, 526–540.

Gresham, F.M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review*, 18, 37–50.

Gresham, F.M. (1996). Treatment integrity in single-subject research. In R. D. Franklin, D. B. Allison, & B.S. Gorman. *Design and analysis of single-case research* (pp. 93–115). New Jersey: Laurence Erlbaum Associates.

Gresham, F. M. (2007). Evolution of the response-to-intervention concept: Empirical foundations and recent developments. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.). *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 10–24). New York: Springer.

Gresham, F. M. (2009). Evolution of the Treatment Integrity Concept: Current Status and Future Directions. *School Psychology Review*, 38, 533-540.

Gresham, F.M., Gansle, K. & Noell, G, Cohen S., & Rosenblum, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980–1990. *School Psychology Review*, 22, 254–272.

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

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

Harcourt, Inc. (1919). Harcourt School Publishers [Core Reading Curriculum]. Retrieved from http://www.harcourt.com/bu_info/harcourt_school.html

Happe, D. (1982). Behavioral intervention: It doesn't do any good in your briefcase. In J. Grimes (Ed.). *Psychological approaches to problems of children and adolescents* (pp. 15–41). Des Moines, IA: Iowa Department of Public Instruction.

Heppner, P.P., Kivlighan, D.M., & Wampold, B.E. (1999). *Research design in counseling*. California: Wadsworth.

Hintze, J. M. & Matthews, W. J. (2004). The generalizability of systematic direct observation across time and setting: A preliminary investigation of the psychometrics of behavioral observation. *School Psychology Review*, 33, 258–270.

Hunter, M. (1982). *Mastery Teaching.* El Segundo, CA: Tip Publications.

Institute for Education Sciences (2009). Request for Applications: Education Research Grants. Retrieved from http://ies.ed.gov/funding/pdf/2009-84305A.pdf.

Kazdin, A. E. (1980). Acceptability of time out from reinforcement procedures for disruptive child behavior. *Behavior Therapy*, 11, 329–344.

Kelley, M.L., Heffer, R.W., Gresham, F.M., & Elliott, S.N. (1989). Development of a modified treatment evaluation inventory. *Journal of Psychopathology & Behavioral Assessment*, 11, 235–247.

Kerlinger, F. N. (1986). *Foundations of behavioral research* (3rd ed.). New York: Holt, Rinehart & Winston.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. NY: Guilford Press.

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. (2nd ed.) NY: Guilford Press.

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

Martens, B. K., Peterson, R. L., Witt, J. C., & Cirone, S. (1986). Teacher perceptions of school-based intervention: Ratings of intervention effectiveness, ease of use, and frequency of use. *Exceptional Children*, 53, 213–223.

Martens, B.K. & Meller, P.J. (1989). Influence of child and classroom characteristics on acceptability of interventions, *Journal of School Psychology*, 27, 237–245.

Martens, B.K., Witt, J.C., Elliott, S.N., & Darveaux, D. (1985). Teacher judgments concerning the acceptability of school-based interventions. *Professional Psychology: Research and Practice*, 16, 191–198.

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

McKee, W. T. (1984). *Acceptability of alternative classroom treatment strategies and factors affecting teachers' ratings*. Unpublished master's thesis, University of British Columbia, Vancouver.

Merrell, K. W. (2003). *Behavioral, social, and emotional assessment of children and adolescents*. New Jersey: Lawrence Erlbaum Associates, Inc.

National Association of School Psychologists. (2005). Position statement on Prevention and Intervention Research in the Schools. Retrieved from http://www.nasponline.org/about_nasp/pp_prevresearch.aspx.

National Association of State Directors of Special Education (NASDE). (2005). *Response to intervention: Policy considerations and implementation*. NASDE: Alexandria, VA.

National Association of State Directors of Special Education (NASDE) (2008). *Response to intervention: Blueprints for implementation school building level*. NASDE: Alexandria, VA.

National Center on Student Progress Monitoring (NCSPM) (2008, October 1). From http://www.studentprogress.org/

National Institute of Child Health and Human Development (NIH) (2000). *Report of the National Reading Panel. Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.

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

Noell, G. H. & Gresham, F. M. (1993). Functional outcome analysis: Do the benefits of consultation and pre-referral intervention justify the costs? *School Psychology Quarterly*, 8, 200–226.

O'Dell, S. L., Tarler-Benlolo, L., & Flynn, J. M. (1979). An instrument to measure knowledge of behavioral principles as applied to children. *Journal of Behavior Therapy and Experimental Psychology*, 10, 29–34.

Pearson Education, Inc. (2008). Aimsweb [Curriculum Based Measurement]. Retrieved from http://www.aimsweb.com/

Peterson, L., Homer, A., & Wonderlich, S. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis*, 15, 477–492.

Popper, K. R. (1994). Science: Problems, aims, responsibilities. In *Myth of the framework: In defence of science and rationality* (pp. 82–111). UK: Routledge.

Reimers, T. M., Wacker, D. P., & Koeppl, G. (1987). Acceptability of behavioral treatments: A review of the literature. *School Psychology Review*, 16, 212–227.

Roberts, G., Good, R., & Corcoran, S. (2005). Story retell: A fluency-based indicator of reading comprehension. *School Psychology Quarterly*, 20, 304–317.

Sattler, J.M. & Dumont, R. (2004). *Assessment of children: WISC-IV and WPPSI-III supplement*. La Mesa, CA: Jerome M. Sattler, Publisher, Inc.

Schilling, S. G., Carlisle, J. F., Scott, S. E., & Zeng, J. (2007). Are fluency measures accurate predictors of reading achievement? *Elementary School Journal,* 107, 429–448.

Schulte, A.C., Easton, J.E., & Parker, J. (2009). Advances in Treatment Integrity Research: Multidisciplinary Perspectives on the Conceptualization, Measurement, and Enhancement of Treatment Integrity. *School Psychology Review*, 38, 460–475.

Shinn, M. R. (2007). Identifying students at risk, monitoring performance, and determining eligibility within response to intervention: Research on educational need and benefit from academic intervention. *School Psychology Review*, 36, 601–617.

Shinn, M. R., Powell-Smith, K. A., Good, R. H., & Baker, S. (1997). The effects of reintegration into general education reading instruction for students with mild disabilities. *Exceptional Children, 64, 59-79.*

Skinner, C. H., Rhymer, K. N., & McDaniel, E. C. (2000). Naturalistic direct observation in educational settings. In E. S. Shapiro & T. R. Kratochwill (Eds.). *Conducting school-based assessments of child and adolescent behavior* (pp. 21–54). New York: Guilford Press.

Sprague, J.R. & Walker, H.M. (2005). *Safe and healthy schools*. New York: Guilford Press.

U.S. Department of Education. (1999). *Twenty-first annual report to Congress on the implementation of the Individuals with Disabilities Act*. Washington, DC: Government Printing Office.

U.S. Department of Education. (2004). U.S. Department of Education, Office of Special Education Programs, Data Analysis System, 2004 IDEA Part B Child Count.

Von Brock, M. B. (1985). *The influence of effectiveness information on teachers' ratings of acceptability*. Unpublished master's thesis, Louisiana State University, Baton Rouge.

Von Brock, M. B. & Elliott, S. N. (1987). The influence of treatment effectiveness information on the acceptability of classroom interventions. *Journal of School Psychology*, 25, 131–144.

Voyager Expanded Learning, Inc. (1994). Voyager Passport and Read Well [Reading Curriculums]. Retrieved from http://www.voyagerlearning.com/about/index.jsp

Waas, G.A. & Anderson, G.P. (1991). Outcome expectancy and treatment acceptability: Perception of school-based interventions. *Professional Psychology: Research and Practice*, 22, 149–154.

Wedl, R. & Schroeder, J. (2005). Response to intervention: An alternative to traditional eligibility criteria for students with disabilities. *Education Evolving*, 1-17.

Wireless Generation, Inc. (2010). mCLASS. [Data Management System]. Retrieved from http://www.wirelessgeneration.com/

Witt, J.C. & Elliott, S. N. (1985). Acceptability of classroom management strategies. In T. R. Kratochwill (ed.). *Advances in school psychology*, 4, 251–288. Hillsdale, NJ: Erlbaum.

Witt, J.C., Elliott, S. N., & Martens, B. K. (1984). Acceptability of behavioral interventions used in classrooms: The influence of teacher time, severity of behavior problems, and type of intervention. *Behavioral Disorders*, 10, 95–104.

Witt, J.C. & Martens, B. (1983). Assessing the acceptability of behavioral interventions used in classrooms. *Psychology in the Schools*, 20, 510–517.

Witt, J.C., Moe, G., Gutkin, T.B., & Andrews, L. (1984). The effect of saying the same thing in different ways: The problem of language and jargon in school-based consultation. *Journal of School Psychology*, 22, 361–367.

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 11, 203–214.

Wright, J. (2005, September). Syracuse NY City Schools. SBIT Resource Packet [Introductory Staff Survey and Referring Teacher Questionnaire]. Retrieved from http://www.jimwrightonline.co/pdfdocs/sbit/sbit_Forms_Sept05/sbitForm6.pdf

Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatment: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology*, 49, 156–167.

Ysseldyke, J., Christenson, S., Pianta, B., & Algozzine, B. (1983). An analysis of teachers' reasons and desired outcomes for students referred for psychoeducational assessment. *Journal of Psychoeducational Assessment*, 1, 73–83.