EXPLORATORY RESPONDER ANALYSIS OF FIRST STEP TO SUCCESS:

STUDENT CHARACTERISTICS CONTRIBUTING TO RESPONSE

by

JEREMY AUSTIN JONES

A DISSERTATION

Presented to the Department of Special Education and Clinical Sciences
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2013

DISSERTATION APPROVAL PAGE

Student: Jeremy Austin Jones

Title: Exploratory Responder Analysis of First Step to Success: Student Characteristics Contributing to Response

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Special Education and Clinical Sciences by:

Dr. Cynthia Anderson	Chairperson
Dr. John Seeley	Core Member
Dr. Hill Walker	Core Member
Dr. Christopher Murray	Core Member
Dr. Lynn Kahle	Institutional Representative

and

Kimberly Andrews Espy	Vice President for Research and Innovation; Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2013

DISSERTATION ABSTRACT

Jeremy Austin Jones

Doctor of Philosophy

Department of Special Education and Clinical Sciences

June 2013

Title: Exploratory Responder Analysis of First Step to Success: Student Characteristics
Contributing to Response

This study examined the ability to identify subgroups of students with unique
characteristics that are related to a meaningful and reliable response to the First Step to
Success intervention. Identifying particular subgroups that are likely to respond or are not
likely to respond to First Step to Success will help practitioners decide when the
intervention should be used with an individual student based on his or her characteristics.
Data on intervention recipients of a large-scale efficacy trial of First Step to Success was
analyzed to identify student characteristics that best predict an initial response to First
Step to Success, as well as the student characteristics that best predict maintenance over
time. Data analysis consisted of a multistep process of examination, including:
determining intervention response, validating the identified responder group,
identification of predictors of initial responders, and identification of predictors of
maintained responders. Results indicate that the only predictor of initial intervention
response was the operant function of a student's problem behavior being attention (OR =
6.67, $\chi^2 = 10.83$, df = 5, $p < .05$. Nagelkerke $R^2 = .22$). No statistically significant
predictor was identified for students who maintained response to the First Step to Success
intervention. Clinical and pedagogical implications of the findings are discussed and
suggested directions for future research are provided.

CURRICULUM VITAE

NAME OF AUTHOR: Jeremy Austin Jones

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene

DEGREES AWARDED:

Doctor of Philosophy, School Psychology, 2013, University of Oregon
Master of Science, Special Education, 2010, University of Oregon
Bachelor of Science, Psychology, Sociology, and Religious Studies, 2004,
University of Oregon

AREAS OF SPECIAL INTEREST:

Parent training
Behavior disorders
Social and emotional development
School-family partnership

PROFESSIONAL EXPERIENCE:

Pre-Doctoral Intern, Implementation Sciences International, Inc., Eugene,
Oregon, 2012-2013

Marte Meo Interventionist, Oregon Social Learning Center, Eugene, Oregon,
2010-2011

Graduate Student Representative and Executive Committee Member, Center on
Early Adolescence, Eugene, Oregon, 2008-2010

Research Assistant, Institute on Violence and Destructive Behavior, Eugene,
Oregon, 2007-Present

Project Coordinator, Oregon Center for Applied Science, Eugene,
Oregon, 2006-2007

Research Assistant II, Oregon Center for Applied Science, Eugene,
Oregon, 2004-2006

GRANTS, AWARDS, AND HONORS:

Helena DeGnath Wessela Memorial Scholarship, University of Oregon, 2011-
2012

Hill M. Walker Legacy Scholarship, University of Oregon, 2010-2011

Helena DeGnath Wessela Memorial Scholarship, University of Oregon, 2010-2011

Technology Entrepreneurship Program Fellowship, University of Oregon, 2010

Ford Foundation Graduate Scholar, Ford Family Foundation, 2008-2010 2011

Ford Foundation Undergraduate Scholar, Ford Family Foundation, 2002-2004 2011

PUBLICATIONS:

Felver, J. C., Doerner, E., Jones, J., Kaye, N. C., & Merrell, K. W. (2013). Mindfulness in school psychology: Applications for intervention and professional practice. *Psychology in the Schools.* Advance online publication. doi: 10.1002/pits.21695

ACKNOWLEDGMENTS

My journey to this point has been in development for a lifetime. I owe much to my parents and the rest of my family, in particular my grandfather, Max Samuel Mintz. Without my grandfather, I would not have realized this path existed and was achievable. Further, the strength of my family and their willingness to sacrifice has allowed me to dedicate myself and pursue this goal. The support and love from my wife Megan has helped to create a stable family and I will always be grateful. The joy of and love from my children, Jada and Keira, have provided me with the nourishment necessary to endure the hardships that this journey has brought. On those seldom occasions where I contemplated giving up my goal, the vision of disappointment upon my family's faces was enough to quickly extinguish the thought and motivate me to continue the pursuit.

I have sincere appreciation for my dissertation committee members, Drs. Cynthia Anderson, John Seeley, Hill Walker, Christopher Murray, and Lynn Kahle for their support and patience throughout this dissertation. In addition, I so very much appreciate all that Dr. John Seeley has provided me. For the past decade, John has provided true mentorship and without him I would not have started this doctoral program or completed it. The world of academia would be significantly better off with more people like John.

There are two other individuals that standout as having imparted an important influence on me during my graduate studies. My previous advisor, the late Dr. Kenneth Merrell, was a tremendous support during my graduate studies and his ability to relate to my life and context was amazingly valuable. Lastly, Dr. Jeffrey Sprague provided mentorship and support of all kinds during these past several years and I am so amazingly appreciative because his support also benefited my family substantially.

This work is dedicated to my teachers. Without them this work would not exist.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

It is estimated that students with severe emotional and behavioral problems comprise between 1% and 5% of the school population, but exhaust more than 50% of teachers' and administrators' time (U.S. Department of Education, 2000). This discrepancy stresses an environment that is continually witnessing declining resources and higher expectations for student achievement. In addition to disrupting the academic learning environment, most students with problem behaviors that do not receive effective intervention and supports will continue to display some degree of problem behavior throughout their lives, exacting a toll on other social institutions (Bullis, Walker, & Stieber, 1998; Patterson, Reid, & Dishion, 1992). Public schools are the de facto service provider for children with problem behaviors (Merrell, Ervin, & Peacock, 2011). Educational institutions need to not only be able to effectively identify students with or at risk for emotional and behavioral disorders, but must also be able to identify and select effective and reliable interventions to ameliorate the problem behavior of those students identified as needing additional supports. A significant challenge that schools face is identifying an individual's unique need and then appropriately matching that particular student with a specific intervention. The challenge is complicated because there is a general lack of research regarding child specific characteristics that are predictive of a response to specific interventions.

Accounting for Individual Differences

While a given intervention may have a substantial evidence base with consistent effects, it does not necessarily mean that when a practitioner utilizes the intervention with

a particular individual it will have the same effect as demonstrated previously. For example, in group analysis, a main effect may be significant because it produces a large effect within a particular subgroup of the analysis, when it may not provide any benefit or demonstrate effectiveness with a different subgroup. Further, the intervention may even cause harm or exacerbate the problem being intervened upon with a particular subgroup. Specifically because of this statistical occurrence, the Society for Prevention Research has outlined standards of evidence for efficacy, effectiveness, and dissemination (Flay et al., 2005) that call for subgroup analyses of evidence-based practices to demonstrate the level of effectiveness of the intervention with particular subgroups. In addition, Kratochwill and Shernoff (2004) have recommended that guidelines be provided with evidence-based interventions on detailing when a particular intervention is and is not likely to be effective to help facilitate the process of schools selecting an evidence-based intervention that will produce predictable, beneficial, and effective results.

An additional impact of applying an evidence-based intervention with a student that is ultimately non-responsive is that it leads to diminished social validity of the intervention by those involved in implementation and ultimately to diminished sustainability of the intervention (Loman, Rodriguez, & Horner, 2010). For example, if a teacher agrees to implement an intervention but the student fails to respond, the teacher will be less likely to implement the intervention again in the future even though it may be highly effective for the next student needing intervention, which may prevent that student from receiving an available and effective intervention.

At some level, choosing one intervention over another for a student exhibiting behavior problems in the early elementary years is somewhat of a guessing game,

especially without knowledge of key factors that may mediate intervention outcomes. The possibility is that either intervention may work as typically delivered, neither may work, or one would work when the other would not. In order to aid in selecting the most effective intervention, there would ideally be variables that are readily identifiable when assessing individuals that are in need of additional supports and that would aid in better predicting the probability of response for a particular individual that has been matched to a particular intervention. When interventions efforts are appropriately matched with individual characteristics and account for individual differences in need they often produce significant outcomes (La Greca, 1993).

Intervention Matching

In schools, best practice in intervention selection entails only using those interventions that are evidence-based, meaning that they are empirically supported and substantiated through quality research that has shown the intervention produces predictable, beneficial, and effective results (Forman & Burke, 2008). Unfortunately, for a variety of reasons, best practices often do not prevail and interventions are often selected that do not have an empirical basis (Ysseldyke, Burns, Dawson, Kelley, & Morrison, 2006). Part of the explanation behind the discrepancy between best practice and actual practice may come down to the complexity of appropriately matching an evidence-based intervention with a particular child. To identify an appropriate intervention practitioners must identify the student with the problem or need, operationalize the problem or need, define intervention goals, and then determine based on the available information what intervention to utilize (Forman & Burke, 2008). In addition to the intervention matching the particular problem or need of the student, a key

element that aids in matching the intervention to the particular student or group of students is often missing. Specifically, what are the characteristics of the individuals that the evidence-based intervention has been effective with in the past? For example, with First Step to Success (First Step; Walker et al., 1998), a commonly used evidence-based intervention, there is little information that describes who would be a good candidate for the intervention other than being identified though a class-wide screening process as having moderate to severe behavior problems and being in kindergarten, first, second, or third grade. Beyond these basic inclusion criteria there is no information demonstrating who it may be most effective with, or potential populations that it has demonstrated to be ineffective with.

What is missing when considering whether First Step might be appropriate are specific criteria that aid in identifying who is likely to respond to the intervention. In other words, why should a school select First Step over some other tier II intervention for a student exhibiting behavior problems? Clearly, if the student was in high school we would rule it out. However, what if the student was a first grader who exhibited a variety of problem behaviors such as talk outs, leaving the classroom without permission, and occasionally spitting on peers? If this described child was to receive First Step what would be the probability that the child would positively respond to the intervention? The answer to the questions are unknown, but it may be possible to acquire certain information prior to selecting an intervention that would allow the school to select an intervention that is truly matched to the individual and one that would reliably predict a response to the intervention.

Study Purpose

        The purpose of this study was to conduct a responder analysis of First Step to better understand student specific variables that contribute to intervention outcomes. The data analysis process detailed in the methodology section assesses student level variables that may significantly influence an individual's response to First Step. Knowledge of student variables that influence intervention response would be beneficial to practitioners attempting to select an appropriate intervention to ameliorate problem behavior in a particular student. The information derived from the analyses will serve to better identify those students who are appropriate candidates for the First Step intervention as well as aid in identifying points where the intervention may be modified in order to better serve those students who are not responding to treatment.

CHAPTER II

LITERATURE REVIEW

This review of the literature focuses on several important topics related to response to intervention, First Step, and predicting treatment outcomes and responder analyses. This section begins by reviewing and discussing the multi-tiered structure of supports and interventions currently being implemented in many schools. This is followed by an in-depth description of First Step including a review of the extant literature. The next section focuses on possible predictors of treatment outcomes and responder analyses in general with a focus on studies conducted within the field of education. The review of the literature concludes by articulating the potential contributions of the study. This review is not intended to be exhaustive, but to provide a foundation for the study.

Multi-tier Prevention

A major effort underway in the education system is adoption and utilization of multi-tier prevention frameworks. The purpose of multi-tier prevention frameworks are to identify at risk students early and provide appropriate services to prevent or ameliorate challenges (Cheney, Flower, & Templeton, 2008; Gersten & Dimino, 2006; Gresham, 2005; McIntosh, Campbell, Carter, & Dickey, 2009).

The general consensus in the field is that multi-tier systems consist of several key principles: a proactive and preventative system that uses school wide screening to identify needs, ensures an appropriate match between student needs and intervention through multi-tiered service delivery, uses a problem-solving orientation and data-based decision making, uses effective evidence-based practices, and is a systems-level approach

(Barnes & Harlacher, 2008; Clements & Kratochwill, 2008; Horner, Sugai, & Anderson, 2010; NASDSE, 2006).

One example of a multi-tiered system is Positive Behavior Interventions and Supports (PBIS). PBIS is a multi-tiered framework consisting of evidence-based interventions and systems to support implementation so as to better enable educators to meet classroom and school-wide disciplinary needs (Sugai, Horner, & Gresham, 2002; Walker, Cheney, Stage, & Blum, 2005). PBIS is a system that looks to transform the school culture as well as the effectiveness of the academic process and research has demonstrated that schools that have implemented PBIS have reduced the number of office discipline referrals (Sugai, Sprague, Horner, & Walker, 2000). Further, once the school culture begins to shift and the atmosphere improves, schools become more orderly and manageable environments (Walker, Ramsey, & Gresham, 2004). PBIS emphasizes prevention of behavior problems within a three-tiered model.

Another example of a multi-tier prevention framework is Response to Intervention (RtI). Gresham (2005) noted four major strengths of the RtI model over the historical "wait to fail" model, which is that it identifies struggling students, it is a model based on risk versus based on deficits, it reduces identification biases, and it focuses on student outcomes. While RtI was popularized by being integrated as policy in the Individuals with Disabilities Education Improvement Act of 2004 (IDEA, 2004), it is not limited to just students with disabilities or as a means of identifying students eligible for special education, it can also be used as an effective model to address behavioral problems (Cheney, Flower, & Templeton, 2008).

Multi-tiered supports. Most multi-tiered models (e.g., PBIS, RtI) have three tiers of progressively more intense levels of intervention supports. At tier I, universal supports and instruction are provided to all students in the school to promote successful student outcomes. Research has demonstrated that schools and school-based practitioners with proper training can provide effective Tier I support that reduces the number of students needing additional and more intensive support (Horner, Sugai, & Anderson, 2010; Scott, 2001). However, even if tier I supports are occurring appropriately, this will only meet the need of approximately 80% of students, with the additional 20% requiring more intensive supports (Horner et al., Sugai, Horner, Gresham, 2002). Further, approximately 15% of students that receive supports at the Tier II level will be responsive to that increased level of support, leaving approximately 5% of students that will require intensive supports at the Tier III level to achieve successful outcomes.

Tier I. Tier I supports are designed to encompass the entire school system and all students should receive exposure to these supports (Greenwood, Kratochwill, & Clements, 2008). A considerable amount of research has been focused on universal Tier I supports and there are a variety of Tier I supports that are utilized in schools today. One of the most common researched and evidence-based interventions for behavior is School-Wide Positive Behavior Interventions and Supports (SWPBIS; Horner & Sugai, 2000), which is a multi-tier system with strong Tier I interventions and support.

Tier I of SWPBIS involves the development of behavioral expectations that then are defined operationally for various school contexts (e.g., the cafeteria). Expectations and the derived rules are explicitly taught to students using behavioral skills training. Compliance with the school expectations and rules are encouraged and reinforced and

appropriate and predefined consequences are implemented for inappropriate behaviors. A considerable amount of research has identified Tier I of SWPBIS as an effective method for decreasing problem behaviors and increasing prosocial behavior and academic success (Bohanon et al., 2006; Leedy, Bates, & Safran, 2004; Metzler, Biglan, Rusby, Sprague, 2001; Warren et al., 2003).

Another universal Tier I system that is being widely adopted in schools in order to enhance student success in school and life is social and emotional learning (SEL) curriculums (Greenberg et al., 2003; Zins and Elias, 2006). Elias et al. (1997) described SEL as a process of acquiring core competencies in order to identify and regulate emotions, set and achieve positive goals, appreciate others perspective, establish and maintain positive relationships, engage in responsible decision making, and handle interpersonal situations constructively. The Collaborative for Academic, Social, and Emotional Learning (CASEL, 2005) have identified the goals of SEL curriculum to foster the development of the five interrelated competencies of self-awareness, self-management, social awareness, relationship skills, and responsible decision making. Applying SEL as a tier I system of support entails providing systematic instruction by teaching, modeling, practicing, and applying the content to the whole school environment so that it is incorporated as part of their daily repertoire of behaviors (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011). Research has demonstrated that providing SEL curriculum helps improve student social behavior, helps students more effectively handle distress, as well as helping students improve academically (Greenberg et al., 2003). Further, Durlak, Weissberg, Dymnicki, Taylor, and Schellinger (2011) conducted a meta-analysis of 213 school-based universal social and emotional learning programs involving

9

270,034 kindergarten and high school students and concluded that SEL programs significantly improve students' skills, attitudes, and behaviors.

Evidence-based Tier I interventions such as SWPBIS and SEL that are effectively implemented at the universal level should provide the necessary supports to enable the majority of students to be successful in the school environment. However, there will still be some students that will need additional more intensive supports offered either at Tier II or Tier III.

Tier III. At the opposite end of the continuum from Tier I are Tier III interventions. Even if schools are effectively implementing Tiers I and II, it is expected that there will be a small percentage of students (approximately 1-5%) that require more intensive intervention and supports in order to be successful. At the Tier III level, instruction is tailored to reflect the specific individual's academic and behavioral goals. Similar, to Tier I interventions, Tier III interventions have received considerable amount of attention from the research community. To match the level of need that student's require, Tier III individualized supports are typically resource intensive and require extensive expertise to develop and often to implement. Tier III social behavioral supports typically rely upon functional behavior assessments to guide the development of the intensive intervention and are therefore specifically tailored to the individual (Fox & Davis, 2005; Gettinger & Stoiber, 2006; McLaren & Nelson, 2009). Because of the severity of need, Tier III supports consist of multi-component interventions designed to prevent problem behaviors, teach desired behavior, and consistently apply consequences for inappropriate behavior in order to decrease the inappropriate behavior (Anderson & Borgmeier, 2010). Ideally, if schools provide all students access to Tier I supports and Tier II (described next) supports are available, then few students will require a tier III intervention.

Tier II. The majority of school-based intervention research has been related to Tier I and Tier III interventions (Anderson & Borgmeier, 2010). Tier II interventions comparatively have seen little research even though they are thought to be required by approximately 15% of the school population in order to be successful. Tier II supports are provided to those students who do not respond to the universal supports provided at Tier I, yet are not considered to need intensive interventions that are typically used at the Tier III level. Once students are identified as needing greater support than what the Tier I universal support provides students will be matched to a Tier II intervention based on perceived student need and what interventions are available at a particular school. Tier II interventions should be easily implemented and ideally will require few extra resources beyond the Tier I universal supports because they are designed to provide efficient supports for a large number of students with similar needs and are therefore most often provided in group settings (Crone, Hawken, & Horner, 2010; McIntosh, Campbell, Carter, & Dickey, 2009). Essential features of Tier II interventions include, explicit instructions of skills, structured prompts for appropriate behavior, opportunities for students to practice newly acquired skills in a natural setting, and frequent feedback to the students, while some may also include mechanisms for fading support and a means of regular communication with the student's parents (Anderson & Borgmeier, 2010). Two of the more common Tier II interventions are Check in-Check out (CICO) and First Step.

Check-in Check-Out is a Tier II intervention designed to be used within a SWPBIS framework. CICO supports students by setting up daily opportunities for acknowledgement from teachers and school personnel as they "check-in" and "check-out" so it is designed for students that are reinforced by adult attention (Crone, Hawken,

Horner, 2010). The key features of CICO are related to increasing student instruction regarding behavioral expectations during the check in and check out process and during teacher feedback, increasing positive interaction with adults during the check in and check out process and with teacher interactions, providing a structured prompt for desired behavior in the form of a daily point card, and lastly frequent feedback for the student on their behavior as well as acknowledgement in the form of check out incentives (Crone, Hawken, Horner, 2010). CICO has a number of advantages that make it a popular Tier II intervention. Specifically it is easily implemented with a large number of students (approximately 30 per individual conducting checks in and out) at one time and can be delivered by a paraprofessional reducing delivery costs, can be continually implemented, has built-in progress monitoring in the form of a daily progress report, and has demonstrated positive effects across multiple studies to date (e.g., Crone, Hawken, Horner, 2010; Fairbanks, Sugai, Guardino, & Lathrop, 2007; Filter et al., 2007; Hawken, MacLeod, & Rawlings, 2007; March & Horner, 2002).

First Step to Success

First Step was developed in the mid 1990's by Walker and colleagues at the University of Oregon, the Oregon Social Learning Center, and the Oregon Research Institute (1997). Since the inception of First Step, it has been extensively researched and utilized in schools across the country (Sumi et al., 2012). The primary goal of First Step is to help children at risk for antisocial behavior appropriately adapt early to the educational environment to promote school success (Walker et al., 1997).

The foundation of First Step is rooted in social learning theory, which identifies social contingencies as key elements in creating behavioral outcomes (Patterson, 1982;

Walker et al., 1997). Specifically, First Step was designed for children who have grown up in a context that unknowingly reinforced antisocial behavior where children have learned to employ efficient, yet coercive antisocial strategies that are then brought to the educational environment when they begin their formal education (Mayer, 1995; Patterson et al., 1992; Walker et al., 1997). Within the educational environment these antisocial strategies typically lead to a whole host of negative consequences, including social isolation and rejection by fellow students and teachers, further solidifying long-term negative outcomes (Patterson et al., 1992; Reid, Patterson, & Snyder, 2002).

Importantly, the First Step approach assumes that the problem solution does not reside within the child, but in the child's social environment. To correct this problem, First Step attempts to reverse this learned antisocial behavior pattern early on in a child's education by changing the social context to replace the existing contingency structure and teach the child more appropriate and equally or more efficient skills and strategies. The First Step interventionist aids in changing the social context by partnering with the child's teacher as well as with the child's caregiver(s) to target the social context where the antisocial behavior pattern most likely is reinforced. As designed, First Step provides access to both adult and peer attention for engagement in appropriate behavior and minimizes access to adult and peer attention for engagement in problem behavior.

There are a number of specific mechanisms underlying First Step that contribute to the success of the intervention. First Step involves explicit instruction in appropriate in-class behavior by the coach explicitly teaching the child appropriate and inappropriate behaviors in the classroom via use of role play and direct instruction. Once appropriate behaviors are taught, a structured prompt is provided for the student in the form of a card

that is red on one side and green on the other side. When the student is displaying appropriate behavior the coach holds the card so that the green side is visible to the student indicating to the student that he or she is demonstrating appropriate behavior that will earn the child points on the card. In addition to this prompt reminding the student to continue maintaining the appropriate behavior and functioning as a stimulus control, it also provides feedback that his or her behavior is appropriate or if the red side is displayed it is inappropriate behavior and he or she is not getting points. Therefore, this red/green card works as a prompt to continue appropriate behavior as well as providing feedback on whether the students behavior is appropriate or not. In addition, the points that are being tallied on the card for appropriate behavior are a type of token economy and function as an additional source of feedback on the student's behavior because a daily reward is only achieved for a certain number of points. Further a typical daily reward in First Step is the student earning free time for his entire class, which also provides peer reinforcement for maintaining appropriate behavior and then the card is sent home where the parents further reinforce their child for appropriate behavior if the minimum number of points are attained. Importantly, First Step allows the student to continually put their skills in to practice because the intervention is conducted within the normal classroom environment.

As demonstrated in a variety of research studies, First Step has ameliorated antisocial behavior in many individuals who have participated in the efficacy and effectiveness studies that have been conducted (e.g., Beard & Sugai, 2004; Carter & Horner, 2007; Carter & Horner, 2009; Diken, Cavkaytar, Batu, Bozkurt, & Kurtyilmaz, 2010; Diken & Rutherford, 2005; Overton, McKenzie, King, & Osborne, 2002; Sprague

14

& Perkins, 2009; Sumi et al., 2012; Walker, Golly, McLane, & Kimmich, 2005; Walker, Kavanagh, Stiller, Golly, Severson, & Feil, 1998; Walker et. al, 2009).

Research studies. Over the past two decades, effects of First Step have been examined in eleven experimental and quasi-experimental studies including nine efficacy trials, both single subject (Beard & Sugai, 2004; Carter & Horner, 2007; Carter & Horner, 2009; Diken & Rutherford, 2005; Sprague & Perkins, 2009) and group design (Diken, Cavkaytar, Batu, Bozkurt, & Kurtyilmaz, 2010; Overton, McKenzie, King, & Osborne, 2002; Walker, Kavanagh, Stiller, Golly, Severson, & Feil, 1998; Walker et. al, 2009), and two effectiveness studies (Sumi et al., 2012; Walker, Golly, McLane, & Kimmich, 2005).

Efficacy studies. Efficacy studies test the response that a specific intervention has on a particular population. Researchers strive to maintain a high quality of implementation of the intervention, which typically involves intensive training of implementers involved in the study, as well as control over confounding factors in order to demonstrate how the intervention when implemented in an ideal manner and context can produce a particular response (Flay et al., 2005). To date, the efficacy of First Step has been demonstrated across nine studies. Studies include both single-subject studies (Beard & Sugai, 2004; Carter & Horner, 2007; Carter & Horner, 2009; Diken & Rutherford, 2005; Sprague & Perkins, 2009) designed to isolate causal variables and document functional relations and group-design studies (Diken, Cavkaytar, Batu, Bozkurt, & Kurtyilmaz, 2010; Overton, McKenzie, King, & Osborne, 2002; Walker, Kavanagh, Stiller, Golly, Severson, & Feil, 1998; Walker et. al, 2009) documenting

15

effects of the standard intervention across larger groups. Several efficacy studies are summarized below.

      Single subject studies. A single subject reversal design study conducted by Beard and Sugai (2004) examined the impact that the parent component has on the intervention in addition to examining overall effects on academic engagement and problem behavior. Six white kindergarten students from two urban elementary schools participated. Three of the students participated in both the home and school components, while the other three only participated in the school-based component. Each of the six participants responded to the intervention as evidenced by increased academic engagement, which averaged approximately 70% prior to intervention and averaged 90% post intervention and decreased problem behavior, which went from an average of one occurrence per observation session to almost zero for each student post intervention. Interestingly, absence of the parent component did not lead to diminished intervention effects on students. Follow-up results at five months post intervention identified four out of the six children as exhibiting a maintained response to the intervention with the other two children requiring additional intervention sessions to maintain initial gains.

      Sprague and Perkins (2009) examined direct and collateral effects of First Step in four kindergarten classrooms using a multiple baseline across participants design. In each of the classrooms, effects of First Step were examined with four individuals; a target student that received the First Step intervention, an in-class peer exhibiting problem behavior, a typical in-class peer, and the classroom teacher. Each of the four target students received First Step across an approximately three month period in a single-subject multiple baseline design across participants. Results of the Sprague and Perkins

16

study demonstrated that First Step had intervention effects for all target students as measured by direct observation of academic engaged time, which increased from an average of 64% at baseline to 90% with First Step implemented and the combined average frequency rate of problem behavior decreased from 4.83 in baseline to 1.20 during intervention. Further, those gains were maintained at follow-up after the intervention had concluded for the target students. Teacher ratings of target student behavior showed significant increases on the social and adaptive measures, as well as significant decreases on maladaptive and aggressive behavior. A unique and important contribution of this study is the effects that First Step had on the other students that were not the First Step target student, as well as the effect on the teacher. The four non-targeted in-class students exhibiting problem behavior also increased their academic engaged time from 69% at baseline to 89% immediately post intervention, as well as moderately increasing adaptive and prosocial behavior while decreasing maladaptive and aggressive behavior over the course of the study without receiving any direct intervention. Sprague and Perkins (2009) speculated that these collateral effects occurring with the non-targeted peers may be attributed to the teacher having more time available to provide positive contact with these peers once the targeted student's behavior improved and potentially more positive reinforcement of their appropriate behavior. In addition the increased modeling of appropriate classroom behavior by the targeted peers may have positively influenced the other non-targeted students with problem behavior. Also of interest, the four typical classroom peers maintained similar status during the course of the intervention for the various measures, meaning they maintained their usual non problematic behavior in the classroom and maintained their same level of academic

engagement. Lastly, the classroom teachers reported an increase in self-efficacy in dealing with student problem behavior as well as improvement of the classroom ecology. This study provides a very unique examination not just of the direct effects of First Step on the children receiving the intervention, but demonstrates the positive impact the intervention can have on the other students in the classroom, the teacher, and more generally the classroom environment.

Diken and Rutherford (2005) studied the outcomes that First Step had with two American Indian kindergartners and two American Indian first graders using a multiple baseline across subjects design. Participant students were put into two groups (kindergarten and first grade). Baseline data were collected simultaneously across groups and once baseline data exhibited stability in play behavior, the first group started the First Step intervention while the second group continued to receive baseline ratings for an initial six days, then started the intervention. The primary outcome of the study focused on the target student's appropriate social play behavior on the school playground. Results of the study demonstrated that First Step significantly improved playground social behavior for the target students with a baseline average of 41% of the observed time spent in appropriate social play, to an increase of 95% during the intervention. The research findings indicate that it is possible that positive reinforcement of appropriate classroom behaviors generalized to the playground environment. However, teacher and parent report of the target students' social behavior were mixed with some recognizing positive improvement in social behavior in line with recorded observational changes and others reporting no change. Further, while the majority of parents and teachers were satisfied with the First Step implementation, one teacher who was unsatisfied with the

18

result of the intervention noted that "cultural factors" likely influenced her perception of the programs limited success, which raises some doubt regarding the cultural appropriateness of the First Step intervention with this population.

Group studies. To date, four group design efficacy studies of First Step have been conducted that document effects of the standard intervention across groups (Diken, Cavkaytar, Batu, Bozkurt, & Kurtyilmaz, 2010; Overton, McKenzie, King, & Osborne, 2002; Walker, Kavanagh, Stiller, Golly, Severson, & Feil, 1998; Walker et. al, 2009). These studies are briefly summarized below.

The first experimental group efficacy study of First Step was conducted by Walker and his colleagues during the 1993 through 1995 school years (Walker, Kavanagh, Stiller, Golly, Severson, & Feil, 1998). A randomized, waitlist control design was utilized. A total of 46 kindergartners (24 in cohort one and 22 in cohort two) with 33 out of 46 measuring in the clinical range on the Achenbach Child Behavior Checklist (CBCL; Achenbach, 1991) participated in the intervention. The majority (74%) of students were male, 7% were of minority status, and 37% were receiving free or reduced lunch. Data was collected at pre and post intervention with follow up data collected through grades 2 for cohort 1 and grade 1 with cohort 2. Study measures consisted of teacher ratings of the students' adaptive and maladaptive behaviors using the Systematic Screening of Behavior Disorders (SSBD; Walker & Severson, 1992), direct observation of academic engaged time, and the Aggression and Withdrawn subscales of the CBCL. Statistically significant effects were found for each of the measures from pre to post (Adaptive, Maladaptive, and Aggression $p < .001$, and academic engaged time $p < .05$)

with the exception of the Withdrawn subscale of the CBCL ($p = .63$). Importantly, these significant effects were still maintained at the one and two year follow-ups.

Overton, McKenzie, King, and Osborne (2002) consulted with and were trained by the original program developers in implementing First Step and conducted a quasi-experimental replication study in five school districts in Oklahoma. A within subjects design with 22 kindergartners who were identified through a universal screening procedure as being at-risk of developing antisocial behavior patterns participated in the study, with a total of 16 who completed the study. The authors examined effects of the intervention on both academic engaged time using direct behavioral observation, and parent and teacher perceptions of child externalizing and aggressive behavior as measured by the CBCL, as well as fidelity of implementation. Study results demonstrated an average improvement in academic engaged time by 31 percentage points for study participants and statistically significant changes from baseline to post intervention on parent ratings of externalizing and aggressive behavior ($p < .05$). However, teacher ratings showed significant improvement only on the Externalizing subscale ($p < .01$) of the CBCL and not the Aggression subscale. Further, one year follow up failed to demonstrate any lasting changes as measured by the parent and teacher CBCL ratings. The authors noted that the median implementation rate for students that successfully progressed through the intervention was 97% and was 71% for those students who did not successfully progress through the intervention. In addition, a number of teachers were inconsistent in implementing core features of the intervention including approximately half being inconsistent with using the red/green card and one-third were inconsistent providing the recommended number of praise statements and minimizing verbal

reprimands. The authors speculated that challenges maintaining fidelity of implementation may be partly responsible for the mixed outcomes of this study.

Walker and colleagues conducted a large scale, randomized controlled efficacy trial, of First Step with 200 students in the Albuquerque New Mexico school district in the 2005 through 2007 school years (Walker et al., 2009). This study was the first large scale randomized examination of First Step with a diverse and large urban school district. The data collected as part of this efficacy trial are the data utilized for this responder analysis and focus on student problem behaviors, direct observation of academic engaged time, adaptive and maladaptive behavior, social skills, and academic performance. The results of this large scale efficacy trial are consistent with previous studies demonstrating statistically significant increases for the majority of students in academic engaged time ($p = .002$), adaptive behavior ($p < .001$), and social skills ($p < .001$), as well as statistically significant decreases in maladaptive ($p < .001$) and problem behavior ($p < .001$). However, it is also important to note that examination of students at the one-year follow up were not statistically significant, meaning the initial gains achieved by the intervention group were no longer significant between the experimental and control groups after one year. An interesting and unique contribution of this study and one that is highly applicable to this current responder analysis is that the researchers linked the fidelity of implementation to the measured outcomes but only partially, leaving large amounts of variance unaccounted for. Therefore, a significant part of the variance may be explained by individual student characteristics.

The first formal experimental test of First Step in a foreign country with a different language and that required translation of the First Step materials was conducted

in Turkey by Diken, Cavkaytar, Batu, Bozkurt, and Kurtyilmaz (2010). A total of 24 students across kindergarten, first, and second grade classrooms from four elementary schools who were identified by teacher rankings of students with problem behaviors participated in the study during the 2007-2008 school year. Half of the students were randomly assigned to the intervention condition while the other half received a treatment as usual condition. Results of the study demonstrated that students receiving the First Step condition were statistically significantly more likely to improve social skills ($p <$ .05) and decrease problem behavior ($p < .05$) compared to the treatment as usual condition as measured by the Turkish version of the Social Skills Rating Scale (Sucuoglu & Ozokcu, 2005). In addition, the majority of teachers and parents participating in the intervention reported satisfaction with First Step, which helps to demonstrate the acceptability and success of the program with diverse populations.

The four group design efficacy studies summarized above provide evidence that First Step has demonstrated effects of the standard intervention across groups. These studies in addition to the single subject studies described earlier have provided evidence of the ability of first step to ameliorate problem behaviors and increase academic engagement in the majority of students with or at risk of antisocial behavior patterns in rather controlled conditions. The next step in establishing the overall effectiveness of an intervention is to conduct effectiveness studies in less controlled conditions.

Effectiveness studies. Effectiveness studies, unlike efficacy studies, test whether interventions are effective under "real-world" conditions (Flay et al., 2005). Two effectiveness studies of First Step have been conducted to date (Sumi et al., 2012; Walker, Golly, McLane, & Kimmich, 2005). Each attempted to examine the effectiveness

of First Step in a manner similar to how an individual school outside a research context would implement the First Step intervention, when there is not direct involvement of the developers or control by external researchers. Effectiveness research is critical to the understanding of how an intervention will provide benefit when implemented in an environment that is not highly controlled or ideal, but similar to the context in which the intervention would be used if adopted on a large scale.

Walker, Golly, McLane, and Kimmich (2005) reported the results of a statewide independent evaluation of First Step conducted by the Human Services Research Institute. In 1999, the Oregon Legislature provided funding to make First Step available to all elementary schools within the state that wanted to implement the intervention. A pre to post quasi-experimental evaluation design was used to determine the effectiveness of First Step with 181 kindergarten through second grade students in Oregon identified by teacher nomination as having social and/or behavioral adjustment problems. A control group was developed by combining the 22 students who comprised a control group from the Walker et al. (1998) efficacy study with data from an additional eight students through the same teacher nomination procedure used to find the intervention participants but did not actually receive the intervention for a total control group of 30 students. Outcome measures were the same measures utilized in the Walker et al. (1998) study. Results replicated previous findings, with students who received First Step significantly increasing academic engagement as measured by direct observation ($p <. 001$) and adaptive behavior as measured by the SSBD ($p <. 001$), while decreasing maladaptive behavior as measured by the SSDB ($p <. 001$) and aggression as measured by the CBCL ($p <. 001$) with pre to post effect sizes of the experimental group ranging from .84 to

23

1.31. In addition, the experimental group's scores differed significantly from those of the control group for all four measures ($p <. 05$). The evaluators determined that teacher satisfaction was moderate to moderately high. Interestingly, fidelity of implementation of First Step ranged from high to very poor, which given the positive results and the fact that this study was designed to mimic real world adoption is somewhat encouraging that the intervention can be effective for the majority of students in less than ideal conditions.

Possibly the most significant study to date is a large scale national effectiveness study reported by Sumi et al. (2012) that involved 48 schools randomly assigned to either the First Step intervention or control condition. A total of 142 intervention students and 144 control students from Illinois, West Virginia, Florida, California, or Oregon, participated in study. Outcome measures were the same as the large scale efficacy study by Walker and colleagues (2009). Results of the effectiveness study (Sumi et al., 2012) are very similar to the efficacy study reported by Walker and colleagues (2009) with students increasing social skills, adaptive behavior, academic engaged time, and decreasing maladaptive and problem behavior. However, while statistically significant effects were found for all the measures with the exception of the direct academic performance measures related to reading ability, effect sizes for the effectiveness studies ranged from .11 to .67 compared to a range of .57 to .87 for the efficacy study. However, this is not surprising and it is a common expectation that effectiveness studies will typically produce smaller effect sizes given that there is less manipulation and control of the context (Flay et al., 2005). Additionally, overall teacher satisfaction was moderate with concerns about the amount of time and resources involved in implementing the intervention as well as detracting from other teaching responsibilities, while parent

satisfaction was favorable. Overall adherence to fidelity of implementation was satisfactory.

The two effectiveness studies examined above have demonstrated that First Step is effective at ameliorating problem behaviors and increasing academic engagement in the majority of students with or at risk of antisocial behavior patterns even under "real-world" conditions when there is not direct involvement of the developers.

A variety of studies, including single subject and group designs, and both experimental and quasi-experimental studies, have provided an evidence base demonstrating that First Step can be implemented with diverse student populations using only those materials and support available to a school as a consumer and still achieve significant results. The current evidence on the positive effects that First Step can produce is substantial. The studies described above have both replicated findings of the previous studies, while also providing unique contributions that help to confirm and describe how First Step works, as well as creating new questions for future research. One of those questions is related to why First Step does not work with all of the students that receive the intervention.

Studies demonstrating limited effect to standardized First Step. Carter and Horner (2007) proposed that a possible reason for a lack of positive response to the First Step intervention may be attributed to the operant function of problem behavior, which First Step as originally designed does not account for. Across two studies (Carter & Horner, 2007, 2009) working with non-responders to First Step, they documented that modifications to First Step based on results of a functional behavior assessment resulted in enhanced outcomes. For example, Carter and Horner (2009) conducted a multiple-

25

baseline across participants design with two kindergartners and one first grader. None of the participants responded to First Step, and each continued to exhibit high levels of problem behavior. Interventions were then modified based on operant function. For example, the intervention options included decreasing access to reinforcers for problem behavior, increasing access to reinforcers for appropriate behavior, and minimizing the effects of triggering antecedents. These modifications resulted in reductions in problem behavior ranging from 37% to 17% relative to baseline. However, it is important to note that teacher and parent ratings of the students' problem behavior and academic engagement did not reflect any significant change, which means that while function-based features may have helped First Step produce a more positive response than the standardized version, it may still have failed to produce a substantial and meaningful response for these particular students. Thus, more research is needed on effects of operant function on First Step outcomes.

Over the past couple of decades a considerable amount of information has been determined by the extensive research conducted on First step. Most importantly, research has indicated that it is an appropriate intervention for many young students with or at risk of antisocial behavior patterns, and that it can effectively ameliorate problem behaviors as well as increase academic engagement for the majority of students. This is true even when fidelity of intervention implementation is less than ideal. However, there are still questions related to the mechanisms underlying the success of First Step, specifically, why the intervention is successful with some students and unsuccessful with others. Certainly, there are limitations with all of the studies described above just as there are with all studies that may limit the effectiveness of the intervention or the ability to

effectively measure response, but a particular question of interest is whether there are unique individual characteristics that play some role in determining the responsiveness of the intervention similar to how Carter and Horner (2009) identified that function based supports tailored to the unique characteristics of the individual can help produce a positive response. In essence, are there individual characteristics that predict a positive response to First Step?

Predicting Treatment Outcomes

There are a variety of reasons as to why an individual may or may not respond to a particular intervention. The majority of research has focused on a few key elements, primarily fidelity of intervention implementation and the amount of intervention received (dose), with limited research focusing on specific characteristics of the individual receiving the intervention. However, if an evidence-based intervention is implemented with fidelity and with a recommended dose and there are still individuals that fail to adequately respond, which does occur with all interventions, this leads one to question why it failed to produce the desired change. What is it about the non-responder individual or the group of individuals that is different from the individuals that did respond to the intervention? Ideally, interventionists would be able to predict treatment outcomes by knowing the characteristics of those individuals likely to respond and those that are not likely to respond. In order to identify characteristics or profiles of individuals that are likely to respond or not respond, a responder or subgroup analysis can be conducted to inform future implementation of specific interventions.

Responder analysis. Responder analyses are very common in the fields of medicine and public health (Kraemer, Frank, & Kupfer, 2006; Petticrew et al., 2012; Sun,

Briel, Walter, & Guyatt, 2010), yet are much less common in the field of education. However, an essential element of an RtI framework is meaningful measurement of response to intervention, meaning there needs to be a clear understanding as to who is responding to evidence-based supports and interventions and who is not responding because it is only with this information that interventionists can modify, maintain, or increase the intensity of the intervention to meet the individual's needs (Gresham, 2004). Therefore, with the increased adoption of RtI in schools, responder analyses have increased. Discussion in the literature has revolved around the difference between statistical significance and meaningful or clinical significance with recognition that the two are not one in the same, with a desire to identify response as meaningful change (Cheney, Flower, Templeton, 2008; Gresham, 2005; O'Connor & Klingner, 2010). In addition, there is increased recognition that even interventions that are evidence-based and implemented with fidelity still fail to produce a significant positive response in many children and that there is little understanding of the variables that predict, influence, or account for the changes and there is a desire for analysis of subgroups to better understand why some children respond and others do not (Eyberg, Nelson, & Boggs, 2008).

One of the primary challenges with conducting a responder analysis is to identify what variables are applicable and should be included in the analysis. When examining individuals there can be many potential reasons as to why they differ from others in the study, so typical analysis entails identifying certain characteristics that the non-responders have in common with each other and yet are different than the responders. It is also suggested that in responder analyses basic demographic variables such as gender,

ethnicity, socioeconomic status, age, as well as those that are applicable to the context of the specific intervention be included in the analysis (Petticrew et al., 2012). In the field of education, responder analyses often evaluate baseline characteristics of the individual related to the construct that is being intervened upon.

Responder analyses in education. It is a natural extension of problem solving and multi-tier frameworks to identify why some students do not respond to evidence-based interventions as this information will better inform intervention selection for a given student. In the field of education there have been relatively few responder analyses conducted to date. Most have focused on reading interventions and suggest that non-responders have more severe verbal or cognitive deficits prior to intervention implementation (e.g., O'Connor & Klinger, 2010; Al Otaiba & Fuchs, 2006; Vaughn et al., 2009). In the area of social behavior, a series of responder analyses have been conducted for CICO (e.g., Cheney, Flower, & Templeton, 2008; Hawken, O'Neill, & MacLeod, 2011; McIntosh, Campbell, Carter, & Dickey, 2009). Some studies related to CICO that analyzed effects of operant function on outcomes suggest that CICO is more effective for students whose behavior is hypothesized to be maintained by adult attention (e.g., Campbell & Anderson, 2008; March & Horner, 2002; McIntosh, Campbell, Carter, & Dickey, 2009) whereas others suggest that function does not mediate outcomes (e.g., Hawken, O'Neill, & MacLeod, 2011).

Responder analyses tend to focus on individual variables that moderate intervention response. However, the majority of research in the field of education related to intervention response focuses on two factors, the fidelity of implementation of an intervention, as well as the intensity or dose of the intervention that the child receives.

The contribution of implementation fidelity to intervention response. Often, lack of intervention effectiveness is attributed to poor fidelity of implementation, which has been documented through numerous studies as accounting for poor intervention outcomes (e.g., Dane & Schneider, 1998; Gottfredson & Gottfredson, 2002; Payne, 2009). That said, previous research by Walker et al (2009) on First Step has demonstrated that fidelity of implementation accounts for only a portion of the variance in outcomes (canonical correlation = .59) leaving a significant portion of the variance unaccounted for. In addition, multiple First Step studies have demonstrated that even when the intervention has been delivered with fidelity, some students still fail to adequately respond (e.g., Carter & Horner, 2007; Carter & Horner, 2009; Walker et al., 2009). Still other studies suggest that, at least in some cases First Step yields positive outcomes even when fidelity is low (e.g., Overton et al., 2002; Sumi et al., 2012). This compilation of data indicates that there are other variables that must be partly responsible for outcomes of the First Step intervention.

The contribution of dose to intervention response. Similar to implementation fidelity, intervention dose has been identified as a contributing factor in intervention response in numerous interventions in both the fields of education (e.g., Elbaum, Vaughn, Hughes, & Moody, 2000; Torgesen et al., 2001; Wanzek & Vaughn, 2008) and psychology (e.g., Barkham et al., 2006; Hansen, Lambert, & Forman, 2002; Harnett, O'Donovan, & Lambert, 2010; McMahon, Wells, & Kotler, 2006). Previous research by Walker et al (2009) on First Step indicates that the amount of dosage of the intervention received accounts for a portion of the variance in outcomes with the canonical correlation for the association between change in outcomes and dosage being $R = .57$. Clearly, this is

a testament to the effectiveness of the actual intervention and a lack of relation would indicate otherwise. However, given that there are instances of students receiving full doses of the intervention yet failing to respond, it is clear that there is still a significant unexplained contribution to intervention response. Part of the unknown contribution to response may be the individual characteristics of the children participating in the intervention.

The contribution of individual characteristics to intervention response. Comparatively little research has examined the role that individual characteristics play in outcomes of interventions for children with behavior problems. However, several characteristics have been identified from previous studies and summarized by McMahon, Wells, and Kotler (2006) in their analysis of predictors of outcomes of interventions with children with behavior problems. Variables identified as moderating previous intervention outcomes for children with behavior problems include: co-occurring internalizing and externalizing symptoms, pretreatment severity of behavior, gender, child age, socioeconomic status, parental status, race and ethnicity, and operant function of behavior.

A challenge with identifying variables that may predict treatment outcomes is that there are often inconsistencies in the research literature. For example, some studies suggest that children who present with co-occurring disruptive behavior and anxiety or depression tend to have less positive outcomes then do children presenting with disruptive behavior only (e.g., Drugli, Larsson, Fossum, & Mørch, 2010), while other research has indicated the opposite having found children with disruptive behavior and co-occurring low levels of anxiety and depression to have a more positive intervention

31

response (e.g., Beauchaine, Webster-Stratton, Reid, 2005). In regards to First Step, there is no published research on whether children with behavior problems and co-occurring internalizing symptoms respond differently than those students with only behavior problems. A possible conclusion is that those children with both externalizing and internalizing behaviors present with a greater severity of behavior prior to intervention compared to those children with just behavior problems, which may play some role in treatment outcomes since those children may need more intensive intervention.

Unfortunately, findings related to the role that pretreatment severity of behavior plays in intervention outcomes are somewhat inconsistent as well with some studies identifying poorer response and others more positive response (e.g., Eyberg et al., 2008; McMahon, Wells, & Kotler, 2006). The majority of studies on First Step have not evaluated severity of behavior as a moderator of intervention outcomes but some research has suggested that severity of pretreatment behavior does not moderate intervention response (e.g., Walker et al., 2009). However, given that First Step was developed for children with mild to moderate behavior problems, children with more severe problems may not be as responsive to the intervention because they are in need of more intensive supports.

Most studies that have assessed and reported on whether gender is a significant moderator of treatment response have not identified any differences between the two genders on intervention outcomes (e.g., Beauchaine, Webster-Stratton, & Reid, 2005; Eyberg et al., 2008). However, there is some limited evidence that school-based interventions designed for prevention of behavior problems are more effective for boys than girls in regards to outcomes (McMahon, Wells, & Kotler, 2006). Most First Step

studies have not reported on gender-specific response rates to the intervention, but Walker et al. (1998) reported no variation in intervention response across genders. It is entirely possibly that boys and girls with behavior problems may respond similarly to an intervention but there is a lack of reporting on this variable leading to difficulty in drawing a conclusion.

Studies have examined child age as a potential moderator of intervention outcomes with some studies identifying younger children as responding more positively (e.g., Dishion & Patterson, 1992; Strain, Young, & Horowitz, 1981) and others finding more positive response for older children (e.g., Bennett & Gibbons, 2000; Serketich & Dumas, 1996) and still others not finding any significant variation in response by age (e.g., Beauchaine, Webster-Stratton, & Reid, 2005; Kazdin, 1995). Given that the target population of First Step is Kindergarten through third grade, there is not a very large age difference. Published research on First Step has not reported on an any significant age specific effects on intervention outcomes, but it does seem possible that older children would be less likely to respond because they have more ingrained patterns of behavior.

Another inconclusive finding relates to socioeconomic status. It is inconclusive whether low socioeconomic status is predictive of a response to intervention given the inconsistent findings related to child and family interventions for problem behavior. Some research reports no significant impact of socioeconomic status on child outcomes and others report a significant negative effect for children living in a household with low socioeconomic status (McMahon, Wells, & Kotler, 2006). Published research related to First Step has not addressed this variable, but it is possible that socioeconomic status could lead to poorer intervention outcomes because the parent component of First Step

could be more challenging for parents with low socioeconomic status and could impact fidelity of implementation.

An additional variable that may differentially impact treatment outcomes is race or ethnicity. McMahon, Wells, and Kotler (2006) describe several studies that have examined ethnicity and race as a moderator variable, but once again, there are inconsistent findings of the effect that race or ethnicity have on intervention response. Additionally, Eyberg, Nelson, and Boggs (2008) report the majority of evidence-based treatment studies for behavior disorders do not include sufficient representation of Hispanic children making this variable difficult to fully analyze. Published First Step research has not reported on differences in intervention outcomes by race or ethnicity, but given that the development of the intervention occurred with a primarily white population, it is possible that important elements of the intervention are not as applicable for minority populations which may impact intervention response.

Lastly, previous research across a number of interventions and populations has indicated that a child's maintaining function of problem behavior (e.g. adult or peer attention maintained or escape maintained) may be a relevant predictor of intervention response (e.g., Ellingson, Miltenberger, Stricker, Galensky, & Garlinghouse, 2000; Hawken, O'Neill, & MacLeod, 2011; Ingram, Lewis-Palmer, & Sugai, 2005; March & Horner, 2002; McIntosh, Campbell, Carter, & Dickey, 2009; Newcomer & Lewis, 2004) including for First Step (e.g., Carter & Horner, 2007; Carter & Horner, 2009). However, there is a need for replicating previous findings related to First Step especially given that the research has only been conducted on a few individuals. It is possible that the operant

function of behavior is partly responsible for intervention response and if this can be

further identified it could inform the intervention process.

In summary, fidelity of implementation, intervention dose, and operant function

have been shown to mediate intervention outcome in at least some cases. At the same

time, there are numerous other variables that either have not been studied sufficiently or

for which existing research has yielded inconsistent outcomes with regard to their

potentially mediating effect. With regard to First Step in particular there are a number of

variables yet to be examined.

Statement of the Problem

There is now a fairly large body of research supporting both the efficacy and

effectiveness of First Step for reducing problem behavior and increasing academic

engagement. Clearly, First Step is successful for many of the children who have

participated in the intervention. However, it is not successful for every child and there is

a lack of understanding as to why it is effective for some children and not others. Fidelity

of implementation and intervention dose are two factors that are often suggested to be a

leading contributors to successful intervention response, however Walker and colleagues

(2009) found that fidelity of implementation and level of dose of First Step only explains

a portion of the variance in outcome, leaving significant amounts unaccounted for.

Therefore it is possible that student characteristics contribute a significant portion.

However, to date, there has been no known statistical subgroup analysis that provides

insights regarding student characteristics that may predict a response to First Step.

Contribution of the Current Study

The current study contributes to previous research on First Step by identifying potential moderating child variables that may predict response to the intervention. To identify student characteristics that best predict an initial response to First Step, as well as the student characteristics that best predict maintenance over time, a multistep process of examination occurred during data analysis. First, an analysis was conducted to group students into two groups, responders or non-responders based on intervention outcomes related to problem behavior scores using the Reliable Change Index (Nunnally & Kotsch, 1983). Second, a validation analysis was conducted to determine if the initial analysis identifying two distinct groups was valid by evaluating whether those identified as responders based on a significant and reliable decrease in problem behavior also significantly changed on other outcome measures commonly associated with responders of First Step. Third, several variables related to student characteristics were evaluated using logistic regression to determine if the characteristics could be identified as predictors of initial responders to First Step. Lastly, the same analysis utilized to identify predictors of initial responders was utilized to identify predictors of maintained responders, or those that maintained significant and reliable declines in problem behavior over a one-year period.

The current study contributes to the field by being the first known study to analyze student characteristics for the purpose of creating an identifiable profile of those likely to respond and not to respond to First Step. Specifically, this study addressed the two following primary research questions:

1. What student characteristics best predict an initial positive response to First Step?

2. What student characteristics best predict maintenance of intervention effects over

   time?

CHAPTER III

METHODS

The data used for this responder analysis was derived from the data collected as part of a large efficacy study conducted by Walker et al. (2009). The participants, procedures, and measures used in the original efficacy study are described in detail, as is the First Step intervention. This is followed by a detailed description of the research methods used for the responder analysis in this study.

Participants

Participants were the First Step intervention recipients in a randomized controlled efficacy trial of the First Step intervention (Walker et al., 2009). The efficacy trial occurred in the 2005-2006 and 2006-2007 school years and included two cohorts of first through third grade students, teachers, and general education classrooms from 34 elementary schools in the Albuquerque Public Schools (APS). Approximately half of the participating schools were involved in this study during the 2005-2006 school year and the remainder in the 2006-2007 school year.

The 200 students who participated in the efficacy study ranged from 6 to 10 years in age at enrollment ($M = 7.2$, $SD = 1.0$) and were predominantly male (73%). Eighty-three students were 1st graders, sixty-nine students were 2nd graders, and forty-eight were 3rd graders. Students were predominantly Hispanic (57%) or Caucasian (24.5%) with the remaining racial and ethnic groups representing less than 20% of the sample (4.5% American Indian, 0.5% Asian or Pacific Islander, 7% Black, 3% multiracial, and 3% unknown). Students came primarily from English-speaking households (88.9%).

Seventy percent were eligible for free or reduced-price lunches and roughly sixteen percent were English language learners.

Given the purpose of identifying child variables that predict response to the First Step intervention, the current analysis included only those participants that received the First Step intervention (not control participants). To examine the sample's representativeness, participating students ($n=200$) were compared to all students who were eligible, but not selected to participate in the study ($n=523$); these students were compared on their baseline demographics and severity of problem behaviors. Participating students were comparable to non-participating students on all demographic variables including age ($M = 7.2$; $SD = 1.0$); t [720] = 1.22), percent female (18.7%; χ^2 [1, N = 723] = 2.97), percent Hispanic (58.5%; χ^2 [1, N = 723] = 0.14), percent Spanish-speaking (13.2%; χ^2 [1, N = 716] = 0.58), percent eligible for free or reduced lunch (77.4%; χ^2 [1, N = 452] = 3.32), and percent ELL (21.5%; χ^2 [1, N = 664] = 2.26). As expected, given that the intervention is designed for and is targeting children with externalizing behavior problems, there were statistically-significant differences (p < .001) between the samples selected for participation and the non-participating samples on the three Systematic Screening for Behavior Disorders (Walker & Severson, 1992) screening measures used to target students with the most severe problem behaviors. Participating students had an average of 6.4 (*SD*=3.8) critical events, an average maladaptive score of 34.5 (*SD*=8.3), and an average adaptive score of 32.4 (*SD*=7.7). In comparison, non-participating students averaged 4.4 (*SD*=3.3) critical events and scores of 30.8 (*SD*=8.4) and 35.3 (*SD*=7.7), respectively, on the maladaptive and adaptive scales.

Measures

Pre and post data were collected with teacher measures, direct observation, and individual academic testing. As part of a larger questionnaire collected prior to and following intervention, teachers completed the grades K-6 version of the Social Skills Rating System (SSRS; Gresham & Elliott, 1990) and a teacher completed functional behavior assessment form. In identifying students for participation in the efficacy study, teachers completed two scales from the Systematic Screening for Behavior Disorders (SSBD; Walker & Severson, 1992). In addition, direct observation data were collected using the SSBD measure of student Academic Engaged Time (AET; Walker & Severson, 1992).

Social Skills Rating System – Teacher (SSRS-T). The SSRS-T (Gresham & Elliot, 1990) is a nationally normed, standardized instrument designed to be completed by teachers of children between the ages of 3-18 years of age and is comprised of 57 items that samples three domains: Social skills, problem behaviors, and academic competence. Each questionnaire takes approximately 10-25 minutes to complete. The test-retest reliability and internal consistency for the teacher SSRS-T scales are .78 to .94 (Gresham & Elliot, 1990).

In the current analysis, only the 18-item Problem Behavior Scale ($\alpha = .85$) was used. The 18 items are equally distributed across three subscales: Externalizing, Internalizing, and Hyperactivity. This scale assesses the teacher's perceived frequency of problem behaviors that may interfere with social skills performance. Each item on the Problem Behavior subscale is assessed on a 3 point scale (never, sometimes, or very often). Total raw scores on the Problem Behavior Scale can range between 0 and 36.

Scores between 0 and 2 (girls) or 0 and 3 (boys) reflect less-than-average levels of problem behavior; scores between 3 and 13 (girls) or 4 and 17 (boys) are considered average levels of problem behavior; and scores between 14 and 36 (girls) or 18 and 36 (boys) are considered to be above-average levels of problem behavior. Raw Problem Behavior Scale scores may be converted to standard scores. There are no standard scores for SSRS-Teacher subscale measures.

Functional behavior assessment. A questionnaire was designed by the original research team to aid teachers in developing a hypothesis about the operant function of each participant's problem behavior. Teachers completed the 25-item questionnaire within the two-week window preceding implementation of First Step with a student. Using the questionnaire teachers provided information about the student's present level of functioning in the school environment, described behaviors of concern and how often they occurred, identified the context in which problems most often occurred, and described consequences that typically followed those behaviors. In addition, teachers completed a summary statement delineating their best guess about events that both evoked and maintained the problem behavior.

Systematic Screening for Behavior Disorders (SSBD). The SSBD, which was used to identify participants for the efficacy study, uses a multiple gating approach to detect students in grades K- 6 who have an elevated risk for school behavior problems (Walker & Severson, 1992). The SSBD consists of three inter-related, and increasingly intensive, screening stages that cross-validate the results of each other. The SSBD screening stages are 1) teacher nomination and rank ordering of up to N (typically 3 to 5) students in a class based on descriptions and examples of externalizing and internalizing

41

behavioral profiles, 2) teacher ratings of students' adaptive and maladaptive behavior and completion of a Critical Events Checklist, and 3) behavioral observations of academic engagement within the classroom and social behavior on the playground. Two behavior-rating scales from Stage Two were completed by teachers prior to and following intervention. The Adaptive Behavior Index (ABI), a twelve-item scale, and the Maladaptive Behavior Index (MBI), an eleven-item scale, assess the student's teacher-related and peer-to-peer adaptive and maladaptive behavioral adjustments based on a 5-point rating scale ranging from "Never" to "Frequently."

The SSBD has excellent psychometric characteristics showing strong evidence for both reliability and validity and is nationally normed (Merrell, 2003; Walker & Severson, 1992; Walker et al., 1990). Specifically, the SSBD has demonstrated strong test-retest reliability over thirty-one days, with the mean test-retest rank order correlations (*rhos*) on the internalizing and externalizing profiles of .74 and .88, respectively (Walker et al. 1990). Further, Pearson correlations for the stage two measures for the Critical Events Index was .81, for the adaptive behavior index was .90, and for the maladaptive behavior index was .87. Internal consistency was calculated for the SSBD national standardization sample resulting in alpha coefficients of .94 for the adaptive behavior rating scale and .92 for the maladaptive behavior rating scale (Walker & Severson, 1992).

Walker et al (1990) evaluated the concurrent validity of the SSBD by examining the correlation between the stage two measures and the Walker-McConnell Scale of Social Competence and School Adjustment (Walker & McConnell, 1988). Correlations between the total score on the Walker-McConnell scale and total scores on the Critical

Events Index, and the Adaptive and Maladaptive Rating Scales were -.57 ($p < .001$), .79 ($p < .001$), and -.44 ($p < .001$), respectively.

Walker et al (1990) also examined discriminant validity of the SSBD from multiple validation and replication studies. Their results demonstrated that each of the three Stage Two measures were effective at identifying students with both externalizing and internalizing problematic behaviors and can discriminate between students with these types of behaviors and those students with normal or well-adjusted school behavior patterns.

Walker et al (1990) also evaluated predictive validity of the SSBD and determined that it is moderately predictive over a one-year period. Specifically, 69% of students that were identified at time one as externalizers were also identified at time two at the one-year follow-up as externalizers. Similarly, 52% of students that were identified at time one as internalizers were also identified at time two at the one year follow-up as internalizers. In addition, 60% of the students that were identified at time one as neither externalizers or internalizers were also identified at time two at the one year follow-up as neither externalizers nor internalizers.

Student Academic Engaged Time. Direct observation data were also collected within two weeks prior to the start of intervention (baseline) and within two weeks after completion of the intervention (post) using the SSBD Stage Three measure of Student AET (Walker & Severson, 1992). Academic Engaged Time estimates, via a stopwatch recording procedure, the amount of time a student spends engaged in allocated academic activities. As described in Walker & Severson (1992), academic engaged time serves as an important indicator of a student's academic involvement and adjustment to the

teacher's classroom expectations for all students. Academic Engaged Time is

operationalized as follows: (a) attending to the material and task, (b) making appropriate

motor responses, (c) asking for assistance at the appropriate time and in an acceptable

manner, (d) interacting with the student's teacher and classmates about academic matters,

and (e) listening to teacher instructions and direction.

A pool of professionally trained observers, blind to student condition, collected

two fifteen-minute AET observations for each student participant at baseline and post

intervention. Observers were recruited from the University of New Mexico and through

advertisements in the local Albuquerque newspaper.

All observers were required to demonstrate and sustain high reliabilities

(minimum .90 inter-observer agreement levels) prior to and during data collection

periods. Observers were monitored, and retrained as necessary, throughout the course of

the study in order to minimize drift and ensure adequate reliability of recorded

observations.

Across the two cohorts and waves of data collection, reliability estimates were

collected on 20% of the recorded AET observations. The intra-class correlation (ICC)

assessing inter-rater reliability for AET observations was excellent (ICC [3, 1] = .99).

For waves 1, 2 and 3 of cohort 1 and waves 1 and 2 of cohort 2, the average AET

reliabilities ranged from .95 to .99 across all baseline and post phases.

These AET assessments were collected on different days within a week of one

another ($M = 2.6$ days, $SD = 2.3$) and were averaged to compute the percent of academic

engaged time for baseline and post time points. For each observation, project staff coded

and recorded the classroom structure (circle time, teacher-led discussion, independent

seatwork, and cooperative learning) and the classroom activity (literacy, math, art or fine motor, and science) in operation at the time of the observation. Most AET observations were collected during circle time (24.8%), teacher-led discussion (31.6%) and individual seatwork (33.8%) activities during which target students were engaged in literacy-related (68.7%) and math-related (19.1%) instructional activities. To minimize effect(s) of varying classroom context(s) on student engagement, every attempt was made to collect post data at the same time of day and during a similar classroom activity and structure as in the pre-intervention observation. There were no statistically significant differences in classroom structure (χ^2 [4, N = 787] = 7.23) or classroom activity type (χ^2 [4, N = 804] = 6.84) between baseline and post data collection occasions.

Fidelity of implementation. Implementation fidelity data were collected by expert raters on four occasions during First Step efficacy trial implementation: once for the behavioral coach during the first five days of program implementation, and on three other occasions, at the beginning, middle, and end of the teacher phase of the program, for the teacher who implemented the program. The Implementation Fidelity Checklist assesses the extent to which the coach and teacher deliver First Step as intended. For each implementation component, the fidelity checklist assesses (a) whether the component was implemented (yes/no), and (b) the quality of implementation using a five-point scale ranging from 0 = Very Poor, .25 = Poor, .50 = Okay, .75 = Good, to 1.0 = Excellent (α = .86). The intra-class correlation assessing inter-rater reliability for implementation fidelity checks was excellent (ICC[3,1] = .92).

Data from the Implementation Fidelity Checklist were used to calculate both adherence and quality implementation scores for the coach, teacher, and overall

classroom. Protocol adherence to First Step implementation was good for both the coach (84%) and teacher phase (82%) of the intervention, with an overall average implementation fidelity percentage of 83%. The quality of implementation averaged .83 for classroom implementation and .76 for the HomeBase components, which indicate mean ratings across intervention components within the good-to-excellent range. With respect to intervention dosage, students received on average 89% of the available classroom program days and 94% of HomeBase sessions. Data related to implementation fidelity indicates that a high rate of fidelity of implementation was achieved during the original efficacy study making this particular dataset ideal for conducting a responder analysis based on individual student characteristics.

Procedures

Efficacy study design. A cohort design model was used in which waves of intervention and usual care comparison students participated in either the 2005-2006 or 2006-2007 school years. In Cohort 1, the first wave of students was identified using the SSBD procedures conducted in the early fall of the 2005-2006 school year. This procedure was replicated for Wave 2 participants in the late fall and then again for Wave 3 participants in the early spring of the 2005-2006 school year. The same procedure were used for identifying participants in the 2006-2007 school year, but there were only two waves recruited in that school year, one in the early fall and another in the late fall. In Cohort 1, there were 99 student participants (44 usual care comparison and 55 intervention); in Cohort 2, there were 101 (55 usual care comparison and 46 intervention).

During the original analysis of the efficacy trial data, potential differences between the two annual recruitment cohorts of students were assessed and no statistically significant interactions with condition were obtained for any of the outcome measures used (Walker et al., 2009). Therefore as in the original analysis, Cohorts 1 and 2 are combined for analysis purposes in this study.

Random assignment. Teachers who consented to participate in the efficacy study ($n = 260$) were randomly assigned to either the intervention or comparison condition and were asked to complete Stages 1 and 2 of the SSBD. Random assignment occurred at the classroom level within waves with only 1 student identified per classroom for study participation. Randomization of identified participants was implemented prior to solicitation of parental consent for their child's participation in the study.

Participant identification. In the SSBD screening Stage 1, teachers were given descriptions and examples of externalizing behaviors and were asked to nominate and rank order the five students in the class who exhibited the highest levels of externalizing problem behaviors. For the three highest-ranked students identified during the first stage, teachers then completed Stage 2 of the SSBD, which included brief ratings of student adaptive and maladaptive behavior and a checklist of 30 high-intensity, low frequency, maladaptive behavioral indicators (e.g., critical behavioral events). The student with the highest average ranking across the SSBD Stage 2 measures was targeted for inclusion in the study.

Externalizing students met criteria for behavior problems in one of two ways using the SSBD scales: (a) if a student had five or more critical events endorsed on the Critical Events Index (CEI) or (b) if a student had one or more (but fewer than five)

critical events endorsed on the CEI, a score of 30 or lower on the Adaptive Behavior Index (ABI), and a score of 35 or higher on the Maladaptive Behavior Index (MBI). Rank ordering of students was based on their CEI, MBI and ABI scores, with the rank order across the three scales averaged for each student. If two students had the same average rank, the one with the higher raw CEI score received the higher rank.

Although every effort was made during the original efficacy study to recruit the first-ranked externalizer within each classroom as identified by the SSBD, on occasion parents of the 1st or even 2nd ranked student declined participation. In total, the SSBD screener was completed on 723 students and the students with the highest average ranking across the SSBD Stage 2 measures were targeted for inclusion in the study, leaving a total of 331 students targeted. Parental consent was attained on 210 students for study participation (63.4%).

For the current analysis, to determine the response of students who received the First Step intervention, only those students that received the intervention were included. Figure 1 below provides an overview of the randomization, screening, and consent procedures for this group.

First Step to Success intervention. The First Step program is a manualized intervention comprised of three modular components: universal screening, classroom intervention (CLASS), and parent training (HomeBase)(Walker et al., 1997). The screening component helps school personnel identify candidates who meet eligibility criteria for program participation. Classroom intervention and parent training comprise the intervention component of First Step.

APS Schools (n=34)

↓

Consented Teachers (n=260)

↓

Randomized to intervention condition (n=125) → Teachers dropped out prior to SSBD screening (n=8)

↓

Teachers (n=117) nominated and rank ordered top 5 externalizing students (n=585)

↓

Teachers (n=117) completed SSBD stage 2 for top 3 externalizing students (n=347)

↓

Target students (n=117)

→ Teachers declined after screening (n=4)

→ Unable to obtain consent with time to complete study

→ All three ranked students declined participation (n=3)

→ Students dropped out after obtaining consent (n=8)

↓

Consented students (n=107)

↓

Participating students at baseline (n=101)

↓

Participating students at post (n=100)
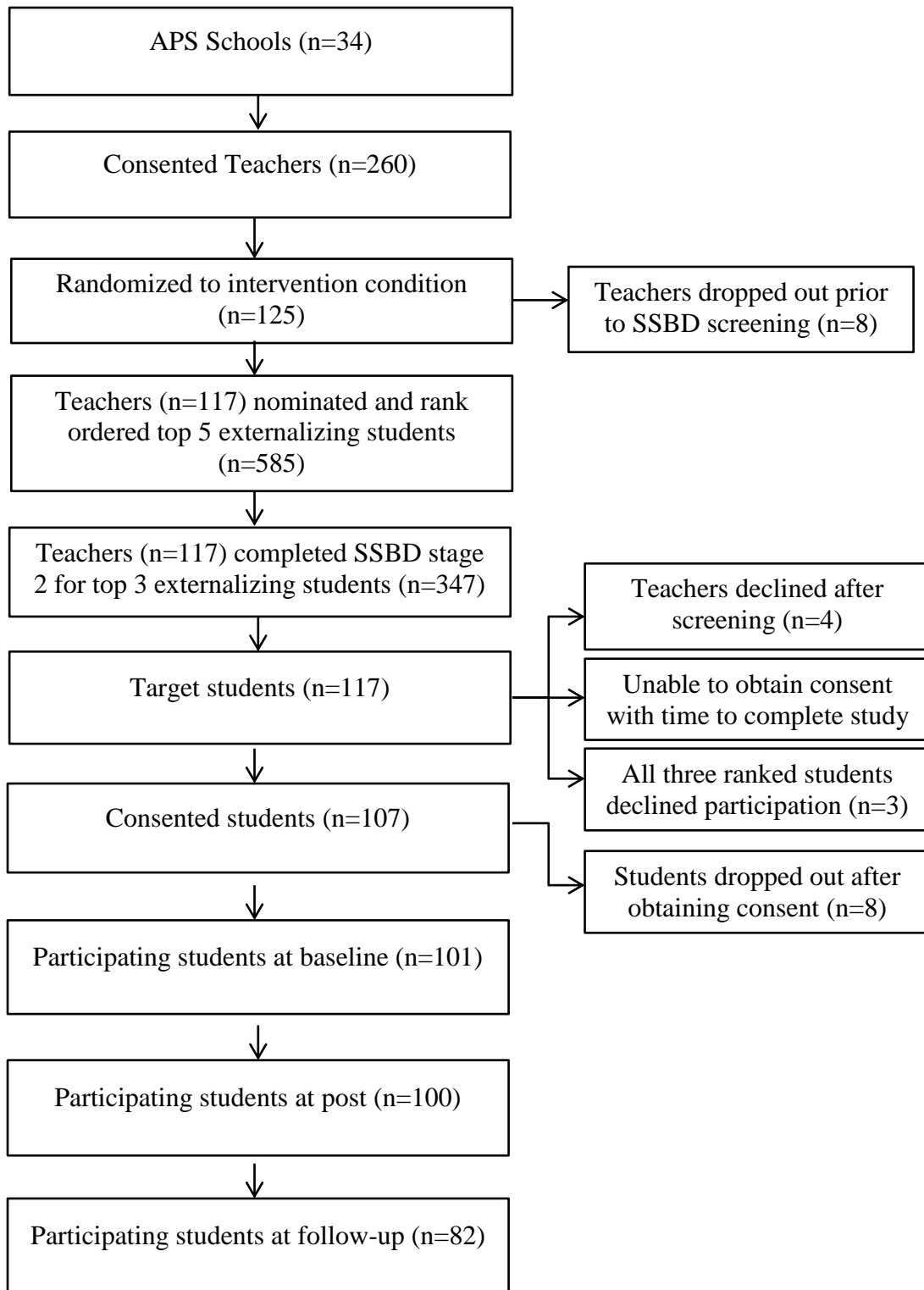
↓

Participating students at follow-up (n=82)

Figure 1. Consort diagram of study participants

During the first five days of the program, a behavioral coach who works with and coordinates the roles of the target child, parent(s), teacher(s), and peers throughout the implementation process, explains and implements the classroom intervention. Typically, the behavioral coach is a counselor, school psychologist, behavioral specialist, or social worker. On the sixth program day, with the support and assistance of the coach, the teacher takes over implementation of the program. On day ten, First Step is extended to the target student's home setting where the coach instructs the parent(s), through six weekly home visits, how to teach their child key school success skills such as communication and sharing, cooperation, problem solving, limit setting, and friendship making. Through instruction, role playing, cueing, prompting, and feedback, parents learn how to teach and encourage these skills in their child.

First Step requires completion of a minimum of 30 program days with a prescribed set of activities, tasks, and a reward criterion. However, prior to the beginning of the intervention with the student and as soon as the child and teacher have been identified, the coach meets with the teacher to discuss general classroom expectations and routines. Ideally, the coach also meets with the parent(s), usually with the teacher, to explain the program, identify parent's concerns regarding their child's behavior in-school and in the home, elicits cooperation with the program, and to identify goals focusing on target behaviors (e.g. following directions). Next, the coach meets with the student to explain the "Green/Red card game", role play appropriate behaviors through examples and non-examples, clarify expectations, and identify possible reinforcers that can be earned for the class and at home. Lastly, the coach conducts observations of the child at

school, typically one 15 minute observation in the classroom and one 15 minute observation during recess to become familiar with the child's behavior.

During the first day of what will ideally be a five day coach led in-class intervention, the coach will meet with the entire class to explain the program and enlist cooperation of the other students and to solicit the help of the target student's peers to discourage engagement with the child when he or she is behaving inappropriately. During this phase the coach implements the program for 20-30 minutes each day. It entails intensive, direct monitoring of the child's behavior using the green/red card, as well as a high rate of consistent positive feedback to the child regarding their appropriate behavior. For example, during the first day of the intervention the student receives feedback every 30 seconds from the coach who is next to the child for the entire 20-30 minutes marking points either on the green side for appropriate behavior or the red side for inappropriate behavior. When the card is on green this signals that the child's behavior is aligned with expectations and that the child is earning points towards a reinforcer. The coach turns the card to red if the child emits a targeted problem behavior; this should serve as a cue to the child to begin emitting desired behavior. Points are only accumulated while the coach or teacher is exhibiting the green side of the card. If the target student exhibits active non-compliance, physical aggression, property destruction, or tantrums during the First Step intervention a time-out routine is implemented, which has already been agreed upon between the teacher and coach and taught to the child prior to the start of the intervention. The coach also provides verbal praise at least nine times on this initial day and the teacher provides verbal praise at least once during this initial day for appropriate behavior.

During this initial coach driven phase, the feedback on the green/red card is quickly faded based on the success of the student achieving the required goal (see appendix 1 for more details). For example, on the 1st day the child is required to achieve 32 out of 40 (80%) possible points on the green side of the card in order to progress to the second day of the program. If they do not receive enough points to progress on any given day (80% of the possible daily points), that day is "recycled" and they must redo that same day in order to achieve the days reward and progress in the program. During each of the five days where the coach is leading the intervention, they are systematically fading their participation while the teacher increases their participation (see appendix 1). In addition, during this initial phase the coach is providing training and support to the teacher to take over the intervention after successful completion of the five program days. During this initial phase, children are typically successful with meeting the criteria to achieve the class or home reward and move to the next day of the program. Importantly, the parents are informed regarding the success of the program at the end of the first day, and will also stay informed regarding the success of the program because the green/red card is sent home every day, which describes if the days goal has been achieved. Parents are instructed to praise and reward for success, and are instructed to only encourage their child to try again the next day if they are unsuccessful.

After the student has achieved success with the five initial program days, the teacher phase begins. At this point the coach has modeled the implementation of the program and provided support and instruction to get the teacher ready to implement independently, so for this phase the teacher takes the lead and is the sole administer of praise and the coach only provides needed support and assistance and helps facilitate

communication between the teacher and parent. The teacher phase is from the 6th to 30th program days. Ideally, the program lasts for only 30 days but this is unlikely given that most children will need to "recycle" days during the program for not achieving the criteria for a particular day just as it happens in the coach phase.

During the teacher phase, the length of the First Step session systematically increases from 30 minutes to 150 minutes, while the intervals between giving praise increases from five to ten minutes. Potential opportunities for points, points needed for the reward, and verbal praise, all gradually raise and then slightly decline, while the displaying of the green/red card consistently declines over the days (see appendix 2 for more details). This systematic process of fading is designed to lead towards generalization of the appropriate behaviors across the classroom day. Eventually, after the end of this phase the green/red card will no longer be used in the classroom and a maintenance plan is established were the teacher utilizes verbal praise to help the student maintain appropriate in-class behaviors.

On program day ten while teacher phase continues in the school, HomeBase, which is the parent component of First Step is started in the student's home environment. During HomeBase the coach instructs the parent(s) through six weekly home visits focusing in on how to help the parent(s) teach their child key school success skills such as communication and sharing, cooperation, problem solving, limit setting, and friendship making.

As part of HomeBase, parents are provided a variety of tools to help set them up for success. All parents receive a Parent Handbook that includes information regarding the skills and activities that the child is learning in the program and "Check-Ups"

53

designed to help the parent identify their child's strength and weaknesses. Also included are Parent Help Cards focusing on specific parenting skills that will help the parent(s) encourage their child to utilize the skills he or she is learning. Parents receive Activity Cards that provide a 10-15 minute activity that the parent(s) and child do together that relates to the skill the child is working on for the week, as well as "Your Ideas" cards that are blank that allow the parent to create activities that will be tailored and enjoyable for their family. In addition, parents receive a timer, used for starting and stopping activities, stickers to be used as rewards, and pen, paper, and markers for various activities.

During the HomeBase phase, the coach will meet with the family for approximately 45-60 minutes per week for six weeks. The goal of the coach is to build a partnership with the parents, offer a positive perspective, help the parent(s) understand the teaching guidelines and to provide support and assistance so that parents may apply the program skills correctly and consistently. In addition, to the weekly meetings, coaches will also contact parents between meetings to help clarify or troubleshoot the week's activities and provide support to the parent(s).

Generally, after the 30 program days of the First Step intervention in a classroom the teacher and coach meet at conclusion of the program and determine whether a maintenance plan should be established. Regardless of whether a specific plan is determined, the coach encourages the teacher to identify the appropriate behaviors that the child is having and praise once every 10 minutes to maintain the child's appropriate behavior. In addition the teacher is encouraged to continue to communicate with the parent(s) regarding their child's progress.

During this particular efficacy trial, the coach and teacher would meet at the end of the 30 program days to tailor a particular maintenance plan. This entailed encouraging the teacher to continue to "catch" the child doing the "right thing", use the phrase "that's green card behavior" or something similar when the child is behaving properly, as well as noticing the target child's peers and providing praise when they are behaving properly. In addition the coach provided a menu of a variety of possible activities, which included activities such as increasing positive verbal feedback, using the green/red card game with the whole class, having a daily or weekly compliment meeting, and continuing to role-play appropriate and inappropriate behaviors, as well as among other similar activities that the teacher could pick from and utilize in the classroom.

First Step coaches. The coaches in this study were drawn from a pool of behavior management specialists and behavior consultants from the APS Behavior Consultation Service Team. A total of 6 consultants and 24 behavior management specialists were trained by a project staff member who was involved in the development of First Step and who had a great deal of experience coaching schools in implementation of First Step. Some elementary schools also expressed an interest in having staff members in their own schools trained as coaches for the school. As a result, there were an additional 18 individuals from Cohort One elementary schools and 3 individuals from Cohort Two who were also trained as coaches. These trainees were certified district personnel, primarily social workers and school counselors, and were trained to the same implementation standards and mastery criteria as coaches from the Behavior Consultation Service Team.

Data Analysis

  To identify student characteristics that best predict an initial response to First Step, as well as the student characteristics that best predict maintenance over time, a multistep process of examination occurred during data analysis, including: determining treatment response, conducting a validation analysis of responder group, identifying predictors of initial responders, and identifying predictors of maintenance responders.

  Measuring response to intervention. In measuring the response to intervention it is necessary to assess the difference between the assessed baseline and post intervention levels of behavior. Gresham (2005) identified five approaches to assess differences between baseline and post intervention phases. These approaches involve examination of data across five metrics: absolute levels of change, the percentage of change between points that meet criterion in the pretest and posttest, percentage of data points that do not overlap between pretest and posttest, effect size, and the reliable change in scores based on standard error. However, currently there is only one known examination of the usefulness of the various metrics suggested by Gresham (Cheney, Flower, & Templeton, 2008), a response to intervention analysis of the Check, Connect, and Expect (Cheney & Lynass, 2009; Cheney et al., 2009) intervention. Given the general lack of adoption and utilization of these various metrics in general and specifically for the First Step intervention it is unclear which if any is the most appropriate method to determine response to intervention.

  When determining the metric to utilize for identifying response to intervention, it is important to examine the applicability of the metric to the type of data involved in the analysis. For example, both the percentage of non-overlapping data points and percentage

of change metrics are designed to be utilized when analyzing frequently repeated measurement of observed behavior and is therefore not applicable given the selected outcome measure of problem behavior as measured by the SSRS.

Gresham (2005) describes three methods of calculating the absolute change metric, the amount of change from baseline to post intervention, the individual no longer meeting criteria for Emotional Disturbance, and the total elimination of behavior problems. Given that it was never a requirement for students to meet criteria for Emotional Disturbance in this study, and there was never an expectation that student's behavior problems would be totally eliminated, the only possible applicable method for this metric would utilizing the amount of change from baseline to post intervention. However, with this current analysis the absolute change metric would only provide either a positive or negative number per individual reflecting whether the individual's problem behavior increased or decreased during the study period. Unfortunately, while it is clear that we would want to see a reduction in problem behavior it is unclear what value would represent significant change.

The effect size metric may be a suitable metric to assess response to intervention for this particular analysis, yet similar to the challenge of using the absolute change metric, there is no agreed upon standard effect size indicative of real behavior change, making interpretation of a particular effect size challenging (Cheney, Flower, & Templeton, 2008).

The advantage that the reliable change index (RCI) has is that it quantifies change from baseline through post intervention. The RCI is calculated by subtracting an individual's post intervention score on the outcome measure from the baseline score and

dividing the difference by the standard error of difference between the post and baseline

scores (Nunnally & Kotsch, 1983). A RCI of ± 1.96 ($p<.05$) is considered reliable change.

This is the same as saying the difference between the two scores has to be at least twice

the standard error of the difference to be significant. Given its ability to quantify change

and determine statistical significance, RCI is the metric utilized to determine whether

significant reduction in problem behavior has occurred between baseline and post

intervention.

Validation analysis. In addition to decreasing problem behavior, an additional

expected result of the First Step intervention was to increase academic engaged time,

which has been demonstrated in previous research findings (Lien-Thorne & Kamps,

2005). It is hypothesized that those students who participated in the study who respond

to the First Step intervention as determined by a significant and reliable decrease in

problem behavior will likewise increase in academic engaged time. If students that

respond to First Step by decreasing problem behavior also increase their academic

engaged time, it will provide further evidence that the response as determine by the RCI

provides an accurate method of assessing response to the First Step intervention. In

essence, this concurrent evidence of response will provide criterion-related validity

indicating that the method proposed to identify accurate response to the First step

intervention is valid.

Criterion validity depends on the ability of a measure to distinguish groups from

one another on an external criterion (Rubin & Babbie, 2001). An analysis of covariance

(ANCOVA) was conducted using group status, either responder or non-responder, which

was identified from the response to intervention analysis as the independent variable,

post-intervention AET scores as the dependent variable, and pre-intervention AET score as the covariate in order to control for baseline scores. If the hypothesis is correct, results will demonstrate that those students who responded to First Step will also see significant increases in AET compared to those students that did not respond. Criterion-related validity will be established if there is a statistically significant difference between responder and non-responders groups on AET.

Predictors of initial response. To address the question of what student characteristics best predict an initial positive response to First Step, a statistical method that can be used for the analysis and prediction of a dichotomous outcome is necessary (i.e. response or non-response). There are a few different statistical methods to address this type of analysis, including ordinary least squares regression, linear discriminant function analysis, and logistic regression. The most appropriate method of analysis to determine if response can be predicted based on baseline characteristics is logistic regression. Further, logistic regression is the most appropriate method to analyze the existing data because the only assumption that will need to be met is the binomial assumption, which is that the binomial distribution is the assumed distribution for the conditional mean of the dichotomous outcome (Peng, Lee, & Ingersoll, 2002).

General guidelines and recommendations by Peng and colleagues (2002) for utilizing logistic regression during analysis and reporting serve as best practice standards for this study. In addition to conducting the logistic regression analysis with contingency tables, evaluation of the logistic model includes an overall model evaluation (e.g., odds ratio and score tests), statistical tests of individual predictors (e.g., Wald chi-square statistic), and goodness-of-fit statistics (e.g., χ^2 test and R^2). An extension of model

validation is also provided where the responder profile is operationalized to determine if an increased response rate occurs based on the existing data.

Predictors of maintained response. Lastly, to determine what predicts maintenance of response of those that initially responded to the First Step intervention, logistic regression was conducted in the same manner as described for predicting the initial response to the First Step intervention.

Predictor variables. In a recent review of evidence-based psychosocial treatments for children with disruptive behaviors, Eyberg, Nelson, and Boggs (2008) noted that "there is little understanding of the variables that predict, influence, or account for the changes in behavior resulting from these interventions" (p. 232). There are also very few applicable guidelines or frameworks as to what variables should be included in a subgroup analysis. Commonly, it is suggested to use basic demographic variables such as gender, ethnicity, socioeconomic status, age, etc., as well as those that are applicable to the context of the specific intervention (Petticrew et al., 2012).

It will be advantageous to identify particular student characteristics known prior to intervention selection that will predict response to intervention so that when an intervention is selected it could be based on data acquired prior to intervention selection, further facilitating data-based decision making. Therefore, in accordance with the primary purpose of this analysis, each of the variables selected at the individual level are identified as predictor variables. Predictor variables will be those that are known prior to implementation of the intervention. This constraint has been chosen because the purpose of this study is to aid school-based practitioners in the identification of which children are

likely to respond and benefit from First Step, and which children are likely to not respond, prior to intervention implementation.

General guidelines detail that selected variables should be specified a priori based on hypotheses about the direction and nature of the effects (Petticrew et al., 2012), and that potential moderators should be baseline measures and should be uncorrelated with the treatment variable (Kraemer, Wilson, Fairburn, & Agras, 2002). There is no identified recommended minimum observation-to-predictor ratio in regards to logistic regression (Peng, So, Stage, & St. John, 2002). However, borrowing from multivariate statistics, it is recommended to adhere to a minimum 10 to 1 ratio, with a minimum sample size of 50. Therefore, the following nine applicable variables were explored in this analysis with the expectation that the overall number of variables would be reduced in the final multivariate model analysis based on statistical and theoretical considerations.

Maintaining function of behavior. Based on previous findings related to First Step and other Tier II interventions (Carter & Horner, 2009; Hawken, O'Neill, & MacLeod, 2011; McIntosh, Campbell, Carter, Dickey, 2009), it is hypothesized that the maintaining function of problem behavior may be a relevant predictor of response to the First Step intervention. Teacher's participating in the original efficacy study completed a questionnaire created for the efficacy trial and designed to identify the maintaining function of the child's behavior. Two raters trained in conducting functional behavior assessments coded the 88 individual assessments as a dichotomous variable, indicating whether the maintaining function is attention maintained or other than attention maintained. An interrater reliability analysis using the Kappa statistic was performed to determine consistency among the two raters who coded the teacher completed functional

61

behavior assessment. The interrater reliability for the raters was found to be Kappa = 0.69, *p* < 0.01, 95% CI [0.50, 0.85]. As rule of thumb, values of Kappa from 0.40 to 0.59 are considered moderate, 0.60 to 0.79 substantial, and 0.80 outstanding (Landis & Koch, 1977). This Kappa indicates that this level of agreement is generally regarded as a reasonably good level of agreement between raters.

Internalizing characteristics. Since externalizing problem behavior was an inclusionary condition of the efficacy trial as determined by the SSBD, all children were considered to exhibit externalizing behavior in the sample. However, there was a subgroup of the sample that had high levels of "internalizing" behavior. The SSRS-T's Internalizing subscale consists of six questions assessing more covert behaviors that reflect social isolation, depressed or sad mood, and anxiousness around other children. Raw scores for this subscale range from 0 to 12. Students that have scores of one standard deviation or higher than the standardization sample comparison group mean on the 6-item internalizing subscale of the SSRS-T can be considered to have above average internalizing behavior symptoms (Gresham & Elliott, 1990).

Children with co-occurring internalizing and externalizing problems typically have higher levels of impairment compared to children with either pure internalizing or externalizing problems (Essex et al., 2009) and higher stability of the symptoms over time (Fanti & Henrich, 2010). Further, children with co-occurring internalizing and externalizing problems are at greater risk than their peers with pure externalizing or internalizing problems for a variety of negative outcomes, including lower academic adjustment and social competence, and higher involvement with antisocial peers. Child co-occurring internalizing and externalizing problems have also been identified as a

pretreatment predictor of externalizing behavior problems at long-term follow-up for children whose parents were participants in a parent training intervention (Drugli, Larsson, Fossum, & Mørch, 2010). Based on the severity and complexity of children with co-occurring internalizing and externalizing behavior problems, it is hypothesized that this subgroup will be less responsive to the First Step intervention in regards to seeing decreased problem behavior.

Severity of behavior. Rankings from the SSBD served as a proxy measure of severity of behavior. There are inconsistent findings in the literature as to whether pretreatment severity of behavior is predictive of response to treatment (Eyberg et al., 2008; McMahon, Wells, & Kotler, 2006). However, it is hypothesized that children with more severe behavior will be more likely to improve behavior and reliably change initially because there is more room for them to change than those with less severe behavior. However, it is also expected that those with more severe behavior will be less likely to maintain improvement of behavior change because of the lack of maintained change in the social environment in which the child initially learned the behavior and in which the behavior is reinforced.

Free or reduced lunch. Free or reduced-lunch eligibility (FRL) will serve as a proxy for socioeconomic status (SES). SES is commonly used as a predictor variable in subgroup analysis (Petticrew et al., 2012). However, it is inconclusive whether low or high SES is predictive of a response to intervention given the inconsistent findings of various studies related to child and family interventions for problem behavior (McMahon, Wells, & Kotler, 2006).

63

Behavior support plan. Students with existing behavior support plans (BSP) would most likely be receiving additional targeted support prior to and during the First Step efficacy study period. Previous findings have determined that multiple or adjunctive interventions typically produce stronger change in behavior over time (McMahon, Wells, & Kotler, 2006). Therefore, it is hypothesized that students with a BSP witnessed greater response during the First Step efficacy trial on key outcomes including greater reduction in problem behavior.

Gender. In general, studies have not identified gender as a significant moderator of treatment response (Beauchaine, Webster-Stratton, & Reid, 2005; Eyberg et al., 2008). However, there is some limited evidence that school-based interventions designed for prevention of behavior problems are more effective for boys than girls in regards to outcomes (McMahon, Wells, & Kotler, 2006). Further, there is limited support for gender as a moderator in long-term outcomes of children with oppositional and conduct problems, with being female predictive of greater oppositional and conduct problems five to six years post intervention. Given that this is a commonly reported and important moderator, as well as recommended variable to use in subgroup analysis (Petticrew et al., 2011), gender was tested as a moderator.

Grade. Studies have examined child age as a potential moderator and have found that the effect on treatment outcomes is inconsistent (McMahon, Wells, & Kotler, 2006). Although age could be examined in this analysis, an examination of grade as a potential moderator would serve a similar purpose with a reduced number of categories and would likely be more informative in the context of First Step.

Ethnicity. McMahon, Wells, and Kotler (2006) describe numerous studies that have examined ethnicity and race as a moderator variable, but once again, there are inconsistent findings of the effect that race or ethnicity may have on intervention response. Eyberg, Nelson, and Boggs (2008) report the majority of evidence-based treatment studies for disruptive disorders do not include sufficient representation of Hispanic children. Therefore, given this studies has ample representation of Hispanic children, a test of the effects of minority status is warranted.

Teacher perception of student's ability to change. A unique variable specific to this dataset is the teacher's perception of the student's ability to change their behavior, which was recorded at baseline. For exploratory purposes, this variable was assessed to determine the ability of teacher's perception to predict response to treatment.

CHAPTER IV

RESULTS AND DISCUSSION

All analyses were conducted using PASW 17.0 Grad Pack (SPSS, 2008). Data on

the primary outcome measure of problem behavior was screened for normality, range

restriction, outliers, and missing data. Data appeared normal with an adequate range and

no severe outliers were identified. A total of 101 students participated in the intervention.

However, one student was missing a problem behavior score at baseline reducing the

overall sample size to 100 students. Missing data appeared to be missing at random. A

description of baseline characteristics for the outcome variable of the SSRS problem

behavior scale and the nine potential predictor variables is provided in Table 1 below.

Table 1

Descriptive Statistics of Outcome and Predictor Variables for Intervention Participants
at Baseline

Variable	Baseline (n=100)
SSRS problem behavior *(M, SD)*	123.02 (1.03)
Teacher-reported internalizing *(M, SD)*	5.00 (0.28)
% Attention maintained PB	60.0
% First grade	40.0
% Female	21.0
% 1st ranked SSBD	84.0
% Current BSP	15.1
% Minority	76.0
% Ability to change	76.0
% FRL eligible	73.3

Identification of Responders

In order to identify a subgroup of students that positively responded to the First Step intervention the RCI was computed for each student. According to the RCI metric, 46 (46.5%) students demonstrated reliable change in reducing problem behavior as evidenced by an RCI greater than -1.96. The average RCI was -1.88, with a range of -9.94 to 3.82. An analysis of variance was conducted to identify any difference between responder groups by baseline problem behavior as measured by the SSRS (Table 2). Results indicate a significant difference between responder and non-responder groups at baseline measurement $F(1, 98) = 646.44$, $p < 0.05$, eta-square = .06, indicating that the responder group has a significantly higher score on the problem behavior scale prior to intervention.

Table 2

Descriptive Statistics for the Social Skills Rating System Problem Behavior Scale by Responder Group at Baseline as Determined by the RCI Metric

	Responders n = 46				Non-Responders n = 53			
Measure	*Min*	*Max*	*M*	*SD*	*Min*	*Max*	*M*	*SD*
SSRS-PB	104	141	125.93	9.75	98	137	120.81	10.16

Note. Min = Minimum reported score; *Max* = Maximum reported score. $F(1, 98) = 646.44$, $p < 0.05$, eta-square = .06.

This significant difference in baseline scores between responder groups may be attributed to inclusion of students in the intervention sample that are already within the normative range on the problem behavior scale. One expectation of First Step is to decrease problem behavior in students with behavior above the normative range, yet it is not necessarily an expectation to significantly change the behavior of students within the

normative range. Therefore, restricting the range of students in the analysis to those with problem behavior above the normative range, which would be those with baseline scores greater than 115 on the SSRS problem behavior scale may be more appropriate in subsequent analyses even though sample size would be reduced.

   Validation analysis of RCI metric. It was previously determined that criterion-related validity will be established if there is a statistically significant difference between responder and non-responder groups on the observational measure of AET. An analysis of variance test was conducted on pre AET scores for responders and non-responders. Results of the analysis indicate that there was not a significant difference between responders (M = 40.70, 95% CI [34.45, 46.94]) and non-responders (M = 44.69, 95% CI [39.29, 50.10]) as measured by AET prior to intervention F(1, 98) = 0.93, p = .33, eta-square = .01. Next, an analysis of covariance was conducted to determine if the responder group that significantly decreased in problem behaviors increased in academic engaged time as originally hypothesized. Results of the analysis indicate that there was not a significant difference between the responder (Madj = 56.71, 95% CI [50.62, 62.80]) and non-responder groups (Madj = 58.24, 95% CI [52.91, 63.57]) post intervention as measured by AET F(1, 97) = 0.14, p = .71, eta-square = .00. Therefore, the non-significant results fail to validate the RCI model that differentiated responder groups. However, this may be an issue with comparing different types of measurement as the observational measure of AET is just a few snapshots in time and may not be capturing the teacher's observation of student behavior and is not necessarily an issue with utilization of the RCI metric. Therefore, an additional validation analysis occurred and is

detailed below utilizing an additional teacher report measure, as well as one that is hypothesized to be related to a decrease in problem behavior.

The social skills scale from the SSRS measure was chosen to evaluate for validation because it is expected that if there is a decrease in problem behavior as demonstrated with those students responding to the intervention, then there should be an increase in social skills. An analysis of variance test was conducted on pre social skills scores for responders and non-responders. Results of the analysis indicate that there was not a significant difference between the responder ($M = 81.81$, 95% CI [79.00, 84.63]) and non-responder ($M = 81.32$, 95% CI [78.92, 83.72]) groups as measured by the SSRS on social skills prior to intervention $F(1, 98) = 0.07$, $p = .79$, eta-square $= 0.00$. Next, an analysis of covariance with responder status as the independent variable and the post-intervention scores from the social skills subscale of the SSRS as the dependent, and the pre-intervention scores for the social skills subscale of the SSRS as the covariate was conducted. Results indicated that students in the responder group ($M_{adj} = 102.63$, 95% CI [98.87, 106.39]) as compared to the non-responder group ($M_{adj} = 85.75$, 95% CI [82.54, 88.95]) significantly increased social skills during the intervention period $F(1, 97) = 46.38$, $p < .01$, eta-square $= 0.39$. Results of this secondary validation analysis provide criterion-related validity and indicate that the RCI metric was adequate for differentiating between responder and non-responder groups.

Variables Predicting Initial Response

Bivariate logistic regression was used to estimate the strength of association of nine variables hypothesized as potential predictors of response to the First Step intervention. Evaluation of all potential predictor variables consisted of assessing for

69

multicollinearity, and evaluating for small or zero cell counts using contingency tables and inflated standard errors in the bivariate analysis. There are no concerns with multicollinearity, however upon assessing cell counts with cross tabulation and conducting initial evaluations of standard errors, grade, ethnicity, SSBD rank, and the variable related to a teacher's perception of a student's ability to change needed to be transformed. Final variables included a dichotomous variable indicating whether the operant function of the student's behavior was attention or other than attention (attention = 1, other = 0), a scale assessing baseline internalizing characteristics of the student (internalizing), a dichotomous variable derived from the SSBD rankings related to the student's severity of behavior (SSBD rank, first ranked =1, second or third = 0), a dichotomous variable indicating eligibility for free or reduced price lunch (FRL eligible = 1, not eligible = 0), a dichotomous variable indicating whether the student had a behavior support plan prior to the start of the First Step intervention (BSP = 1, no BSP = 0), a dichotomous variable indicating whether the student is in first grade or if they are in second or third (Grade, first grade = 1, second or third = 0), a dichotomous variable indicating whether the student is male of female (sex, female = 1, male = 0), a dichotomous variable indicating whether student records list the student as a member of an ethnic group other than white or Caucasian. (Ethnicity, minority = 1, non-minority = 0), and lastly a dichotomous variable related to the teachers perception of the child's ability to change his or her behavior (ability to change, likely = 1, not likely = 0). A bivariate logistic regression analysis (Table 3) was conducted for each of the predictor variables being considered for the multivariate model. Inclusion in the subsequent multivariate models is contingent on statistical significance ($p < 0.25$) as variables

70

meeting this cut-off point may still influence the effect of other variables, as well as

variables that are theoretically or clinically important (Hosmer & Lemeshow, 2000).

Table 3

Bivariate Logistic Regression Analysis of Predictor Variables

Predictor variable	N	OR	95% CI	p
Sex	99	**0.65**	(0.24-1.74)	0.24
Grade	99	**1.78**	(0.79-4.01)	0.16
FRL eligible	89	**0.54**	(0.21-1.40)	0.20
SSBD rank	99	1.14	(0.39-3.35)	0.81
Ethnicity	99	0.67	(0.26-1.67)	0.39
Internalizing	99	**1.24**	(1.06-1.45)	0.01
Ability to change	99	0.93	(0.36-2.37)	0.88
Attention	89	**6.73**	(2.38-19.01)	<0.01
Record of BSP	92	**0.44**	(0.13-1.53)	0.20

Note. Odds ratios shown in bold for variables with $p < .25$

As identified in Table 3, six variables have been identified as significant at the $p <$

0.25 level, including two variables at the $p < 0.05$ level. Given the significant results from

the bivariate logistic regression and the baseline difference in problem behavior scores

between the responder groups, additional analyses were conducted restricting the

intervention sample with the expectation that restricting the range will lead to a non-

significant result between responder groups on baseline problem behavior, yet still

significant results on the logistic regression of predictor variables.

Clinical Sample Responders

A subsample consisting of students above the normative range at baseline on the SSRS problem behavior scale (≥115)was identified reducing the intervention sample to 77 students. An additional benefit of using this sample was that with statistically significant results the analysis could also demonstrate meaningful clinical change because restricting the RCI metric to those students that started above the normative range but to be considered a responder must also decrease in symptomatology to a degree consistent with students within the normative range (<115) is indicative of meaningful clinical change. Specifically, these clinical responder students would decline in problem behavior symptomatology to a degree that could not be attributed to chance or measurement error and back within a normative range making the child virtually indistinguishable from other normally functioning peers in regards to problem behavior. Utilization of this clinical sample also minimizes any issues that may arise when using difference scores (Streiner & Norman, 2008). Therefore, the RCI was recalculated for each student that had a problem behavior score at or above 115 (+1 SD above the normative range) at baseline and below 115 post intervention. According to the RCI metric, 33 (43%) students demonstrated reliable change in reducing problem behavior as evidenced by an RCI greater than -1.96. The average RCI was -2.18, with a range of -9.94 to 2.49. An analysis of variance was conducted to identify any difference between groups by baseline problem behavior (Table 4), which was non-significant $F(1, 76) = 76.58$, $p = .18$, eta-square = .02.

Table 4

Descriptive Statistics for the Social Skills Rating System Problem Behavior Scale by Clinical Responder Group at Baseline as Determined by the RCI Metric

| | Clinical Responders | | | | Non-Responders | | | |
| | *n* = 33 | | | | *n* = 44 | | | |
Measure	*Min*	*Max*	*M*	*SD*	*Min*	*Max*	*M*	*SD*
SSRS-PB	117	139	128.70	6.77	115	141	126.68	6.35

Note. Min = Minimum reported score; *Max* = Maximum reported score.

As expected, the restricted range in the clinical subsample produced a non-significant result at baseline problem behavior measurement unlike when the sample was not restricted to students with above average behavior problems.

Variables Predicting Initial Clinical Responders

Bivariate logistic regression was used to estimate the strength of association of the same nine variables ran in the previous analysis of potential predictors of response to the First Step intervention. Inclusion in the multivariate model was contingent on statistical significance ($p < 0.25$) as variables meeting this cut-off point may still influence the effect of other variables, as well as variables that are theoretically or clinically important such as demographic variables (Hosmer & Lemeshow, 2000).

As displayed in the table 5, three variables met the predetermined criteria of statistical significance at $p < 0.25$. Further, while not statistically significant there are a few variables that are theoretically important and commonly recommended to include in responder analyses (Petticrew et al., 2012). Specifically, FRL eligibility was retained as this is a proxy for socio-economic status and ethnicity was also retained as both are important demographic variables. In addition, BSP was retained in the multivariate model because it is theoretically important because children with preexisting behavior support

plans should be receiving additional services, which may contribute to reducing problem behavior, or on the other hand, these students may stand out more than other peers and have therefore already come to the attention of school staff indicating possible severity of their behavior.

Table 5

Bivariate Logistic Regression Analysis of Predictor Variables for Clinical Responders

Predictor variable	*N*	OR	95% CI	*P*
Sex	77	**0.43**	(0.14-1.35)	0.15
Grade	77	1.17	(0.47-2.93)	0.74
FRL eligible	73	1.19	(0.42-3.37)	0.74
SSBD rank	77	0.88	(0.25-3.19)	0.85
Ethnicity	77	1.17	(0.42-3.30)	0.76
Internalizing	77	**1.26**	(1.04-1.54)	0.02
Ability to change	77	0.78	(0.28-2.22)	0.65
Attention	69	**4.68**	(1.48-14.77)	<0.01
Record of BSP	70	0.50	(0.14-1.77)	0.65

Note. Odds ratios shown in bold for variables with *p* < .25

This leaves three variables that were removed from the analysis. The variable indicating severity of behavior was dropped as it was statistically insignificant and all children in the analysis were already above the normative range of problem behavior as measured by the SSRS reducing the importance of this variable. The variable indicating teacher perception of the student's ability to change was dropped because it was not statistically significant and was not directly a student characteristic. Lastly, grade was not included in the multivariate analysis because it was not significant and has a restricted

74

range given that it was necessary to create a dichotomous variable across three grades because of subgroup size and this also serves somewhat as a proxy for age which overlaps across grades.

Multivariate logistic regression of initial predictor variables. Multiple multivariate models were considered to determine which model was most parsimonious but still explained the data. Unfortunately, given the limited sample size it was possible to overfit the model with too many variables leading to unrealistically large estimated coefficients or estimated standard errors (Hosmer & Lemeshow, 2000). However, there was a desire to retain variables that are not just included because of statistical significance but to also evaluate variables such as demographic variables that when evaluated together may have a confounding effect.

The initial multivariate model includes only those variables that met the criteria for statistical significance (Table 6). Predictors with a bivariate association with problem behavior outcomes with $p < 0.25$ were entered into the multivariate logistic regression model. A total of 69 students were included in the analysis because of missing data due to listwise deletion. Predictors with $p > .05$ were eliminated using backward stepwise logistic regression.

Interpretation of the predictor variables was guided by the likelihood ratio test χ^2 test, Wald statistic, and odds ratio. The Wald statistic tests the hypothesis that the odds ratio is 1, or equivalently, that the regression coefficient is 0. When Wald is statistically significant, this hypothesis is rejected and the odds ratio is judged to differ significantly from 1. The best predictor in this model of a student characteristic of who is likely to respond to the First Step intervention was if a student's operant function of their behavior

Table 6

Logistic Regression Analysis of Statistically Significant Predictors for Initial Multivariate Model

Predictor	Variables in the model at Step 2					
	β	*SE* β	*df*	*Wald*	*p*	OR(CI)
Internalizing	0.21	0.11	1	3.50	0.06	1.23(0.99-1.54)
Attention	1.43	0.60	1	5.69	0.02*	4.16(1.28-13.50)

Note. Likelihood ratio test χ^2 at Step 2 = 11.67, $df = 2$, $p < .01$. Nagelkerke $R^2 = .21$. Hosmer and Lemeshow Test $\chi^2 = 2.79$, $df = 7$, $p = .90$

was attention (OR = 4.16, $p < .05$). This means that a student whose problem behavior is maintained by attention is slightly more than four times as likely to respond to the intervention versus a student whose problem behavior is maintained by another function (e.g., escape, tangible). Neither, internalizing characteristics or sex was identified as a significant predictor, yet internalizing characteristics was approaching significance in this model. The Hosmer-Lemeshow test was insignificant ($p > .05$) suggesting that the model fit the data.

The second multivariate model includes the same variables from the initial multivariate model but this model also forced demographic variables and variables of clinical importance that were previously identified as potential predictors yet did not achieve the predetermined statistical significance criteria to be included in the initial model (Table 7). Variables include ethnicity, free or reduced lunch status, behavior support plan, sex, internalizing characteristics, and operant function of behavior. A total of 61 students were included in the analysis because listwise deletion excludes data with missing values. All variables were forced besides the variables related to internalizing

characteristics and operant function of behavior, which can be eliminated using backward

stepwise logistic regression with $p > .05$.

Table 7

Logistic Regression Analysis of Statistically Significant Predictors for Secondary Multivariate Model

Predictor	Variables in the model at Step 2					
	β	*SE* β	*Df*	*Wald*	*p*	OR(CI)
Attention	1.90	0.77	1	6.06	0.01*	6.67(1.47-30.19)
FRL eligible	0.50	0.75	1	0.45	0.50	1.65(0.38-7.17)
BSP	-0.70	0.90	1	0.60	0.44	0.50(0.08-2.93)
Ethnicity	-0.27	0.78	1	0.12	0.73	0.77(0.17-3.56)
Sex	-0.50	0.72	1	0.49	0.49	0.61(0.15-2.48)

Note. Likelihood ratio test χ^2 at Step 2 = 10.83, $df = 5$, $p < .05$. Nagelkerke $R^2 = .22$. Hosmer and Lemeshow Test $\chi^2 = 3.260$, $df = 6$, $p = .78$

Interpretation of the predictor variables was guided by the likelihood ratio test χ^2

test, Wald statistic, and odds ratio. Again in this model the best predictor of a student

characteristic that indicates who is likely to respond to the First Step intervention was if a

student's operant function of their behavior was attention (OR = 6.67, $p < .05$). This

means that a student whose problem behavior is maintained by attention is slightly more

than six and half times as likely to respond to the intervention versus a student whose

problem behavior is maintained by another function (e.g. escape, tangible) when the other

demographic and clinically important variables are held constant. No other variables were

identified as a significant predictor. However, by forcing the other variables in this

model, they have slightly enhanced the effect of attention maintained behavior as a

predictor. This second model slightly improves upon the initial model $X^2 =10.83$, df = 5,

$p < .05$, Nagelkerke $R^2 = .22$. Further, the Hosmer-Lemeshow test was insignificant ($p >$

.05) suggesting that the model fit the data.

Classification tables for the two models of predictor variables are provided

below in Tables 8 and 9. Predicted probabilities of at-risk status were retained from the

logistic regression analyses. Dichotomized probabilities (below or above .50) were used

to place individuals in their most likely category. The table provides information

regarding the validity of the predicted probabilities using a cutoff set at 0.50. The

percentage listed in the first row of the classification table indicates the magnitude of

specificity or proportion of correctly non-identified students (i.e., non-responder). The

percentage listed in the second row indicates the sensitivity or proportion of correctly

identified students (i.e., responder). Again the second model (Table 9) improved upon

the initial model with overall classification accuracy for the predictor variables in the

logistic regression analysis of 72.1, which is a considerable improvement to the chance

level.

Table 8

The Observed and Predicted Frequencies for Responder Groups in the Initial Model

	Predicted		
Observed	Non-responder	Clinical-responder	% Correct
Non-responder	25	13	65.8
Clinical responder	8	23	74.2
Overall % correct			69.6

Note. Cutoff = 0.50.

Table 9

The Observed and Predicted Frequencies for Responder Groups in the Second Model

	Predicted		
Observed	Non-responder	Clinical-responder	% Correct
Non-responder	27	9	75.0
Clinical responder	8	17	68.0
Overall % correct			72.1

Note. Cutoff = 0.50.

After evaluating multiple logistic regression models using statistically significant and clinically meaningful potential predictors, the second model was considered the final model because it clearly identified a statistically significant predictor of response to the First Step intervention and the model was statistically sound with significant results and good fit to the available data.

Operationalized Clinical Responder Profile

In order to extend model validation of the final multivariate logistic regression model the clinical responder profile was operationalized to determine if an increased response rate occurs on the existing data (Table 10). All students that had a problem behavior score above the normative range, which would be indicative of typical students receiving the First Step intervention in a school environment, were analyzed to determine response dependent on whether the operant function of their behavior was attention or not attention.

Table 10

The Percent of Students with Clinically Meaningful Response by Operant Function Type

| Responder group | Operant Function of Behavior | | Total | Rate by Function |
	Not attention	Attention		
Non-responder	17	20	37	46%
Clinical responder	5	26	31	84%
Total	22	46	68	
Overall response	22.7%	56.5%	45.6%	

The results indicate that for students whose operant function of their problem behavior was attention, 56.5% of those students responded by both decreasing problem behavior scores in a statistically significant degree, as well as in a clinically meaningful way by decreasing problem behavior scores significantly enough to move within the normative range. This is a 10.9% increase over the total population included in the analysis and is considerably higher than the 22.7% response rate for students whose operant function was identified as something other than attention.

A clear profile utilizing predictor variables based on student characteristics has been identified with the only predictor demonstrating significance immediately following completion of the First Step intervention is attention as the operant function of the problem behavior. The next step of the analysis was to see if there is also a profile identifiable for students maintaining response to the First Step intervention over a one year period.

Study Attrition

Prior to examining predictors of maintained response at one-year follow-up an examination of attrition of the intervention participants was warranted. Results indicate that while problem behavior scores were recorded for 100 participants at baseline, only 82 students had ratings of problem behavior at follow-up, which is a decrease of 18 students. The clinical responder classification variable was examined to determine if attrition was discrepant between responders and non-responders, A relatively equal percentage of responders (21.2%) and non-responders (20.5%) were missing problem behavior data at follow-up. Further, as indicated below (Table 11), there are no particular characteristics that were more likely to contribute to attrition. Therefore, any attrition appears to be at random and other than decreasing sample size attrition should not prove to be impactful of the generalizability of the results.

Table 11

Attrition from Baseline to Follow-up by Student-level Baseline Characteristics

Characteristic	Baseline (n=100)	Follow-up (n=82)	Test statistic	*p*-value
Teacher-reported internalizing (*M, SD*)	5.00 (2.82)	5.12 (2.82)	0.32	.57
Attention maintained PB % (*n*)	60.0 (60)	58.5 (48)	0.03	.87
First grade % (*n*)	40.0 (40)	39.0 (32)	0.47	.50
Female % (*n*)	21.0 (21)	20.7 (17)	.09	.76
1st ranked SSBD % (*n*)	84.0 (84)	80.5 (66)	3.92	.05
Current BSP % (*n*)	15.1 (14)	13.2 (10)	1.17	.28
Minority % (*n*)	76.0 (76)	78.0 (64)	1.21	.27
Ability to change % (*n*)	76.0 (76)	75.6 (62)	0.42	.52
FRL eligible % (*n*)	73.3 (66)	73.3 (55)	0.02	.88

Note. Reported test statistics are *t* for continuous and χ^2 for dichotomous measures.

Variables Predicting Maintained Response

To determine whether students who initially responded to the First Step intervention maintained their response over time until the follow-up assessment, all responders were evaluated to determine if follow-up measurement of problem behavior was still within the normative range. A total of only 12 of the initial clinical responders maintained response, which is approximately 36% of those that demonstrated an initial clinical response.

Bivariate logistic regression was used to estimate the strength of association of the same nine variables run in the previous analysis of potential initial predictors of response to the First Step intervention. Inclusion in the subsequent multivariate model was contingent on statistical significance ($p < 0.25$).

As displayed in Table 12, two variables met the predetermined criteria of statistical significance at $p < 0.25$. Additionally, the variable indicating that a student had a behavior support plan prior to participating in the First Step intervention was unable to be analyzed because the smaller sample size of maintained responders at follow-up lead to a structural zero cell in the analysis leaving the variable unusable for analysis.

Multivariate logistic regression of maintained predictor variables. The initial multivariate model (Table 13) included only those variables that met the criteria for statistical significance in the bivariate analysis. Predictors with a bivariate association with problem behavior outcomes with $p < 0.25$ were entered into the multivariate logistic regression model. A total of 58 students were included in the analysis because listwise deletion excludes data with missing values. Predictors with $p > .05$ were eliminated using backward stepwise logistic regression.

82

Table 12

Bivariate Logistic Regression Analysis of Predictor Variables for Maintained Responders

Predictor variable	*N*	OR	95% CI	*P*
Sex	65	**0.28**	(0.33-2.38)	0.24
Grade	65	0.71	(0.19-2.63)	0.60
FRL eligible	63	2.19	(0.42-11.25)	0.35
SSBD rank	65	0.53	(0.12-2.41)	0.41
Ethnicity	65	0.98	(0.23-4.15)	0.97
Internalizing	65	1.16	(0.33-4.12)	0.82
Ability to change	65	0.59	(0.15-2.29)	0.44
Attention	58	**7.07**	(0.84-59.57)	0.07

Note. Odds ratios shown in bold for variables with $p < .25$

Table 13

Logistic Regression Analysis of Statistically Significant Predictors of Maintained Response

	Variables in the model at Step 2					
Predictor	β	*SE* β	*df*	*Wald*	*p*	OR(CI)
Attention	1.96	1.09	1	3.24	0.07	7.07(0.84-59.57)

Note. Likelihood ratio test χ^2 at Step 2 = 4.90, $df = 1$, $p < .05$. Nagelkerke $R^2 = .13$. Hosmer and Lemeshow Test $\chi^2 = 0.60$, $df = 2$, $p = .97$

Unfortunately, as identified in Table 13, there are no statistically significant predictors in this initial multivariate model at the $p < .05$ level. Given how the clinically meaningful variables forced in to the final multivariate model used to identify the initial predictor profile enhanced the effect of the lone predictor variable, those same variables were forced in to a subsequent model. However, unlike the multivariate model for the

initial predictor, adding the clinically meaningful variables created a model that failed the likelihood ratio test (X^2 at Step 2 = 7.87, df = 4, p = .10), indicating predictor variables had no increased effects in predicting a maintained response. Therefore, the final model was the initial multivariate model where the odds ratio for attention was comparable to the initial responder analysis albeit the result was not statistically significant (p = .07) due to the loss of power associated with student attrition and decrease in response rate.

CHAPTER V

GENERAL DISCUSSION

The purpose of this study was to analyze student characteristics for the purpose of creating an identifiable profile of those likely to respond and not to respond to First Step, both initially upon completion of the intervention, as well as maintaining response over time until a one-year follow-up assessment. Utilizing previously collected data from a large-scale efficacy study of First Step, a unique metric that utilized the RCI in combination with a clinically meaningful change in problem behavior symptomatology was operationalized to identify responders from non-responders. Second, the responder group was validated, indicating that the metric used in the analysis was a reasonable means of distinguishing responders from non-responders. Third, an analysis of predictor variables known prior to intervention implementation identified a single variable related to a student's operant function of problem behavior as predictive of initial response to the intervention. In addition, the analysis indicated that a variety of other variables including standard demographic variables were not predictive of response, meaning First Step can be successfully utilized with a variety of students. Lastly, an attempt was made to identify variables predictive of maintained response to the intervention, but no variable was a significant predictor of maintained response.

Limitations

It is important to recognize that there are several limitations to this study. First, this was an exploratory analysis. Before decisions are made based upon the findings in this study additional analyses using other datasets is warranted to determine if the

findings can be confirmed. Therefore, caution is urged when interpreting conclusions from this exploratory study.

The second limitation relates to the choice of utilizing the RCI metric. As other researchers have mentioned, the RCI metric tends to be rather conservative (Cheney, Flower, & Templeton, 2008). This means that this study may be underestimating true responders to the First Step intervention. In addition, the results are tied to a particular definition of responsiveness and different definitions of responsiveness may yield different numbers and different responders.

The third limitation relates to the sample size in this study. This exploratory analysis utilized a relatively small sample and missing data and attrition also contributed to reductions in the sample used for analysis. A specific challenge with this small sample size appeared when attempting to analyze maintained responders. There were very few (n=12) maintained responders left in the analysis at follow-up making examination of individual variables challenging. Specifically, the variable indicating which students had behavior support plans prior to intervention implementation could not be included in the analysis because of the lack of sample size.

The fourth limitation of this study pertains to utilizing indirect measurement of the student's operant function of problem behavior with a teacher completed functional behavior assessment form. While there was reasonable inter-rater reliability on the coding of the form (Kappa = 0.69, $p < 0.01$, 95% CI [0.50, 0.85]), the use of this practice is questionable given that the teachers were not trained in conducting this type of assessment and functional behavior assessments typically involve more than a simple completion of a form to identify the operant function of the student's problem behavior.

86

Further, a recent meta-analytic review of the use of indirect assessment methods for functional behavior assessment found that rating scales and other indirect measures are lacking in regards to rigorous analysis of their utility and in general have poor reliability (Floyd, Phaneuf, & Wilczynski, 2005). It is possible that this simplistic process, which makes a lot of sense from a feasibility perspective, introduced measurement error in to the analysis by misidentification of the operant function of behavior for particular individuals.

Discussion of Findings

Any given Tier II intervention should not be expected to work with any given student. There are a myriad of factors that would impact response to the intervention, including whether the intervention addresses the reason the problem is occurring. For example, a skills training intervention will likely be more effective with students exhibiting a skill deficit relative to students for whom contingency management is the issue. Further, the match between intervention characteristics and student characteristics may impact outcomes such that some interventions are more effective with certain students and/or problems. Historically, in school-based practice when an intervention is implemented but not effective, the response often has been to consider eligibility for special education and to begin the process of individualizing an intervention and/or identifying a more restrictive setting. An important first question however is whether the intervention attempted was in fact (a) implemented with fidelity, and (b) a good match for the student.

If operant function of behavior is a moderator to response to First Step as identified in this exploratory analysis, then this is an important variable to have

87

knowledge about. Indeed, it makes sense that this variable may moderate outcomes and that student's whose problem behavior is maintained by attention would be more responsive to the intervention than those whose problem behavior is maintained by another function given that the First Step intervention focuses on changing the social context by replacing the existing contingency structure and providing access to both adult and peer attention for engagement in appropriate behavior and minimizing access to adult and peer attention for engagement in problem behavior. Therefore, if function matters, and it seemingly does, we need to have a means of appropriately addressing this variable in practice.

On way of increasing responsiveness to intervention is to select an intervention that is appropriately matched to the student and his or her needs. A proactive approach based on the findings from this analysis is to conduct a functional behavior assessment prior to intervention selection and if the student's operant function of behavior is attention, then First Step would be an obvious choice given the considerably higher response rate (56.5%) for those students than those whose problem behavior was maintained by something other than attention (22.7%). Unfortunately in practice, a functional behavior assessment is rarely conducted prior to Tier II intervention selection, primarily because this tends to be a resource intensive process. However, if future research is able to determine that a pragmatic teacher rating of perceived motivation (similar to what was utilized in this study) is determined to be a sufficient indicator of operant function of behavior, this typically labor intensive process may be simplified.

A secondary approach to increasing responsiveness to intervention would be to evaluate students who are non-responsive after receiving First Step with a functional

behavior assessment. Those that are identified as attention maintained and non-responders may be good candidates for a more intensive intervention and those whose problem behavior is maintained by something other than attention should probably either receive a different Tier II intervention matched to their needs or to incorporate function-based supports in to First Step to better target and ameliorate the behavior. Regardless of the approach, if these exploratory findings are supported by additional research it seems prudent to evaluate the match of the intervention with the student before increasing intervention intensity or evaluating for or placing a student in special education.

An additional and unfortunate conclusion of this analysis is that a relatively small number of students (n=12) maintained response at the one year follow-up assessment. It is possible that it may be necessary to provide specific supports over time to sustain effects in this population. Therefore, the problem behavior would be treated more like a chronic illness such as diabetes that requires ongoing maintenance versus a surgical excision of problem behavior that would just heal and solve the issue requiring no further involvement. Moreover, considering that the lone variable that predicts response to the intervention is related to attention, it would make sense that students have acquired a strategy that is highly effective, which is engage in problem behavior and receive attention. First Step provides an effective teacher led strategy to replace the student's acquired technique that is still reliant upon providing attention, yet for appropriate behavior. Therefore, if implementation of First Step is removed and the student is no longer receiving attention for appropriate behavior, then why would they not revert back to a strategy that has proven successful in the past, especially if this strategy is reinforced in either the school or home environment or potentially both? A possible solution is to

provide ongoing support, yet this is somewhat contradictory to Tier II interventions that are designed to be short term in nature. Another possibility would be to provide some form of booster session periodically to maintain effects. Regardless of the strategy chosen it seems that having a maintenance plan may be necessary and ideally future research will identify effective strategies that can be incorporated into the intervention.

Findings from this exploratory analysis are akin to findings from previous single subject analyses on First Step indicating that students with attention maintained problem behaviors are more likely to respond to the First Step intervention as typically delivered in accordance with the manual than student's whose behavior is maintained by a different function (Carter & Horner, 2007, 2009). However, it is important to note, that the findings by Carter and Horner also identified that function-based supports could be added to the First Step intervention to enhance the effectiveness of those students that were initially non-responsive and as discussed below this is a potential area of future research. Research on Check-in, Check-out has also concluded that students are more likely to respond to the intervention if their problem behavior was maintained by adult attention (e.g., Campbell & Anderson, 2008; March & Horner, 2002; McIntosh, Campbell, Carter, & Dickey, 2009). Clearly, there is a growing body of research suggesting that function of behavior may moderate response to intervention and additional research on First Step as well as other interventions designed to ameliorate problem behaviors is warranted.

Future Directions

This is an exploratory analysis and additional research should be conducted on First Step to confirm whether the results from this analysis, that attention maintained problem behavior moderates initial response to the intervention can be confirmed. In

addition, analysis of interventions other than First Step should examine whether there are student characteristics, specifically operant function of behavior that are predictive of response.

There very well may be benefit in producing significant and clinically meaningful change even if that change is only temporary. However, the ideal would be to have a lasting significant benefit as this is also a requirement of an evidence-based practice (Flay, et al., 2005). Future research should examine whether there are predictors of maintained responders, as well as evaluate mechanisms of maintenance in order to determine how to enhance long term positive outcomes. One important area of research that warrants examination is related to the role that teachers can play in sustaining intervention outcomes. Specifically, intervention effects may be enhanced if the teacher implementing the intervention knew what was maintaining the problem behavior and maintained awareness of this during interaction. Further, maintenance of response may be more successful if the student's teacher the following year knew for example that attention was maintaining this particular student's problem behavior and keeping this in mind and focusing intervention efforts on this particular element could help generalize the intervention across time and settings. An additional area of research that warrants future research is how parenting behavior and the home context can aid in enhancing long-term outcomes. Indeed, the home context is often where problem behavior develops and may function as a mechanism maintaining problem behavior. Efforts to improve homeBase, the parent component of First Step may prove to be critical in maintaining improvements in problem behavior over time.

If these exploratory finding are further confirmed, additional research examining how First Step can be modified to support response in student's whose operant function of problem behavior is different than attention is also needed. Previous research (Carter &Horner, 2009) has demonstrated that function-based supports can be integrated within the First Step intervention and future research should attempt to replicate these findings, as well as evaluate how to integrate these practices in to the manualized intervention protocol in a manner that is feasible for large scale implementation.

As RtI is implemented on an ever increasing scale, it will be necessary to refine what exactly qualifies as response in behavioral interventions. Greater examination of the appropriateness of the RCI metric is needed because it is possible that this metric is overly conservative as has been discussed in previous research (Cheney, Flower, & Templeton, 2008). An evaluation of this metric in comparison to other metrics described by Gresham (2005) would help provide insight regarding the appropriateness of the metric for measuring response in students receiving the First Step intervention. Identification of an appropriate metric that can clearly identify meaningful and statistically significant change will be important to the field to be able to confidently and accurately identify students who are responsive to an intervention and whether a more intense or just a different type that is a better match is needed to ameliorate the student's symptoms.

Lastly, it will be beneficial for future research to examine how effectively teacher completed assessments can accurately identify a student's functional behavior. Currently, standard practice in schools entails implementing a Tier II intervention prior to functional behavior assessment. This practice is pervasive because of the considerable resources and

expertise required to effectively conduct an assessment of this type. This barrier could be much less burdensome to overcome if researchers identified easy and efficient strategies to identify operant function. In this responder analysis a teacher completed form was able to distinguish with reasonable reliability whether a student's operant function of behavior was attention or not, but to determine truly how reliable, future research should evaluate the accuracy and feasibility of utilizing a teacher completed form in comparison to a gold standard FBA procedure. With information derived from and FBA those tasked with matching an intervention to the student's needs could be properly informed and either select an intervention that is empirically supported to meet that student's needs or modify an existing intervention by accounting for the student's operant function of his or her behavior.

Conclusions

The primary contribution of this study was identifying a profile based on student characteristics that positively responds to the First Step to Success intervention. This profile consists of a single measurable variable that can be obtained prior to intervention implementation. The only variable that was predictive of response to intervention was the operant function of a student's problem behavior, with attention maintained behavior predictive of a statistically significant and clinically meaningful initial response to the First Step intervention. Therefore, upon confirming these findings, it may be possible to significantly increase the initial response rate to the First Step intervention by identifying those students whose problem behavior is maintained by attention. Unfortunately, a small number of students maintained their positive response to the intervention and at follow-up there were not significant predictors of maintained response. Future research should

attempt to confirm the findings of this exploratory study as well as examine ways to

support maintenance of the First Step intervention over time.

APPENDIX

DAILY SUMMARY CHART

Program Day	Length of Session	Maximum Interval between points/positive feedback	Total Points Possible	Points Needed	Verbal Praise C=Coach T=Teacher		Reward Earned
Coach Phase (10-second point opportunities)							
1	20 min.	30 Sec.	40	32	C: 9 T: 1		Daily
2	20 min.	1 min.	20	16	C: 7 T: 2		
3	20 min.	2 min.	10	8	C: 4 T: 3		
4	20 min.	4 min.	5	4	C: 3 T: 4		
5	30 min.	5 min.	6	5	C: 3 T: 4		
Teacher Phase							
6	30 min.	5 min.	6	5	6		Daily
7	40 min.	5 min.	8	6	8		
8	60 min.	5 min.	12	10	10		
9	90 min.	5 min.	18	15	15		
10	120 min.	5 min.	25	20	20		
11-12	120 min.	8 min.	15	12	20		2nd Day
13-15	150 min.	8 min.	20	16	15		3rd Day
16-20	150 min.	8 min.	20	16	15		5th Day
21-30	150 min.	10 min.	15	12	15		10th Day
31-end	Maintenance	10 min.	none	none	15		Occa-sional

Note: 80% or more of the total available points have to be rewarded on GREEN side in order to earn the group activity reward at school.

REFERENCES CITED

Al Otaiba, S., & Fuchs, D. (2006). Who are the children for whom best practices in reading are ineffective? An experimental and longitudinal study. *Journal of Learning Disabilities, 39,* 414– 431.

Anderson, C., & Borgmeier, C. (2010). Tier II interventions within the framework of School-Wide Positive Behavior Support: Essential features for design, implementation, and maintenance. *Behavior Analysis in Practice, 3*, 33-45.

Achenbach, T. (1991). *Child Behavior Checklist/4-18*. Burlington: University of Vermont.

Barkham, M., Connell, J., William, S., Miles, J., Margison, F., Evans, C, & Mellor-Clark, J. (2006). Dose-effect relations and responsive regulation of treatment duration: The good enough level. *Journal of Consulting and Clinical Psychology, 74*, 160-167.

Barnes, A., & Harlacher, J. (2008). Clearing the confusion: Response-to-Intervention as a set of principles. *Education and Treatment of Children, 31*, 417-431.

Beard, K. Y., & Sugai, G. (2004). First Step to Success: An early intervention for elementary children at risk for antisocial behavior. *Behavioral Disorders, 29*, 396-409.

Beauchaine, T. P., Webster-Stratton, C., & Reid, M. J. (2005). Mediators, moderators, and predictors of 1-year outcomes among children treated for early-onset conduct problems: A latent growth curve analysis. *Journal of Consulting and Clinical Psychology, 73*, 371-388.

Bennett, D. S., & Gibbons, T. A. (2000). Efficacy of child cognitive-behavioral interventions for antisocial behavior: A Meta-analysis. *Child and Family Behavior Therapy, 22,* 1-15.

Bohanon, H., Fenning, P., Carney, K., Minnis-Kim, M., Anderson-Harriss, S., Mortoz, K… & Pigott, T. (2006). Schoolwide application of positive behavior support in an urban high school: A case study. *Journal of Positive Behavior Interventions, 8,* 131-145.

Bullis, M., Walker, H. M., & Stieber, S. (1998). The influence of peer and educational variables on arrest status among at-risk males. *Journal of Emotional and Behavioral Disorders, 6*, 141-152.

Campbell, A., & Anderson, C. M. (2008). Enhancing effects of check-in/check-out with function-based support. *Behavioral Disorders, 33,* 233-245.

Carter, D. R., & Horner, R. H. (2007). Adding functional behavioral assessment to First Step to Success: A case study. *Journal of Positive Behavior Interventions, 9*, 229-238.

Carter, D. R., & Horner, R. H. (2009). Adding function-based behavioral support to First Step to Success: Integrating individualized and manualized practices. *Journal of Positive Behavior Interventions, 11*, 22-34. doi: 10.1177/1098300708319125

Cheney, D., Flower, A., & Templeton, T. (2008). Applying response to intervention metrics in the social domain for students at risk of developing emotional or behavioral disorders. *Journal of Special Education, 42*, 108-126.

Cheney, D., & Lynass, L. (2009). *Check, Connect, and Expect: Program manual*: Seattle, WA: University of Washington.

Cheney, D., Lynass, L., Flower, A., Waugh, M., Iwaszuk, W., Mielenz, C., & Hawken, L. (2009). The Check, Connect, and Expect program: A targeted, tier 2 intervention in the schoolwide positive behavior support model. *Preventing School Failure: Alternative Education for Children and Youth, 54*, 152-158.

Clements, M. R., & Kratochwill, T. R. (2008). Multitiered prevention models: Implications and future perspectives. In C. Greenwood, T. Kratochwill, & M. Clements (Eds.), *Schoolwide prevention models: Lessons learned in elementary schools* (pp. 269–289). New York: Guilford.

Collaborative for Academic, Social, and Emotional Learning [CASEL]. (2005). *Safe and sound: An educational leader's guide to evidence-based social and emotional learning programs - Illinois edition*. Retrieved October, 9, 2012, from http://www.casel.org.

Crone, D., Horner, R., & Hawken, L. S. (2010*). Responding to problem behavior in schools: The Behavior Education Program* (2nd edition). New York, New York: Guilford.

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

Diken, I. H., Cavkaytar, A., Batu, E. S., Bozkurt, F., & Kurtyilmaz, Y. (2010). First Step to Success - a school/home intervention program for preventing problem behaviors in young children: Examining the effectiveness and social validity in Turkey. *Emotional & Behavioural Difficulties, 15*, 207-221. doi: 10.1080/13632752.2010.497660

Diken, I. H., & Rutherford, R. B. (2005). First Step to Success early intervention program: A study of effectiveness with Native-American children. *Education & Treatment of Children (ETC), 28*, 444-465.

Dishion, T. J., & Patterson G. R. (1992). Age effects in parent training outcome. *Behavior Therapy, 23,* 719-729.

Drugli, M. B., Larsson, B., Fossum, S., & Mørch, W. T. (2010). Five-to six-year outcome and its prediction for children with odd/cd treated with parent training. *Journal of Child Psychology and Psychiatry, 51*, 559-566.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82,* 405–432.

Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at-risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology*, *92*, 605-619.

Elias, M. J., Zins, J. E., Weissberg, R. P., Frey, K. S., Greenberg, M. T., Haynes, N. M., et al. (1997). *Promoting social and emotional learning: Guidelines for educators.* Alexandria, VA: Association for Supervision and Curriculum Development.

Ellingson, S. A., Miltenberger, R. G., Stricker, J., Galensky, T L., & Garlinghouse, M. (2000). Functional assessment and intervention for challenging behaviors in the classroom by general classroom teachers. *Journal of Positive Behavior Interventions, 2,* 85-97.

Essex, M. J., Kraemer, H. C., Slattery, M. J., Burk, L. R., Boyce, W. T., Woodward, H. R., & Kupfer, D. J. (2009). Screening for childhood mental health problems: Outcomes and early identification. *Journal of child psychology and psychiatry, and allied disciplines, 50*, 562-570. doi: 10.1111/j.1469-7610.2008.02015.x

Eyberg, S. M., Nelson, M. M., & Boggs, S. R. (2008). Evidence-based psychosocial treatments for children and adolescents with disruptive behavior. *Journal of Clinical Child & Adolescent Psychology, 37*, 215-237.

Fairbanks, S., Sugai, G., Guardino, D., & Lathrop, M. (2007). Response to intervention: Examining classroom behavior support in second grade. *Exceptional Children, 73*, 288-310.

Fanti, K. A., & Henrich, C. C. (2010). Trajectories of pure and co-occurring internalizing and externalizing problems from age 2 to age 12: Findings from the national institute of child health and human development study of early child care. *Developmental Psychology, 46*, 1159-1175.

Filter, K., Benedict, E. A., Horner, R. H., Todd, A. W., & Watson, J. (2007). Check-in/Check out: A post hoc evaluation of an efficient secondary level intervention for reducing problem behaviors in schools. *Education and Treatment of Children, 30,* 69-84.

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., . . . Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6*, 151-175. doi: DOI 10.1007/s11121-005-5553-y

Floyd, R. G., Phaneuf, R. L., & Wilczynski, S. M. (2005). Measurement properties of indirect assessment methods for functional behavioral assessment: A Review of research. *School Psychology Review, 34,* 58-73.

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

Fox, J., & Davis, C. (2005). Functional behavior assessment in schools: Current research findings and future directions. *Journal of Behavioral Education, 14,* 1-4.

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

Gardner, F., Connell, A., Trentacosta, C. J., Shaw, D. S., Dishion, T. J., & Wilson, M. N. (2009). Moderators of outcome in a brief family-centered intervention for preventing early problem behavior. *Journal of Consulting and Clinical Psychology, 77*, 543-553.

Gersten, R., & Dimino, J. (2006). RTI (Response to Intervention): Rethinking special education for students with reading difficulties (yet again). *Reading Research Quarterly, 41,* 99–108.

Gettinger, M., & Stoiber, K. (2006). Functional assessment, collaboration, and evidence-based treatment: Analysis of a team approach for addressing challenging behaviors in young children. *Journal of School Psychology, 44,* 231-252.

Gottfredson, D.C., & Gottfredson, G.D. (2002). Quality of school-based prevention programs: Results from a national survey. *Journal of Research in Crime and Delinquency, 39,* 3-35.

Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional and academic learning. *American Psychologist*, 58, 466-474.

Gresham, F. M. (2004). Current status and future directions of school-based behavioral interventions. *School Psychology Review, 33,* 326–343.

Gresham, F. M. (2005). Response to intervention: An alternative means of identifying students as emotionally disturbed. *Education and Treatment of Children, 28*, 328.

Gresham, F. M., & Elliott, S. N. (1990). *Social Skills Rating System*.Circle Pines, MN: American Guidance Service.

Hansen, N., Lambert, M., & Forman, E. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice, 9*, 329-343.

Harnett, P., O'Donovan, A., & Lambert, M. (2010). The dose response relationship in psychotherapy: Implications for social policy. *Clinical Psychologist, 14*, 39-44.

Hawken, L. S., MacLeod, S. K., & Rawlings, L. (2007). Effects of the behavior education program" (BEP) on office discipline referrals of elementary school students. *Journal of Positive Behavior Interventions, 9,* 94-101.

Hawken, L. S., O'Neill, R. E., & MacLeod, K. S. (2011). An investigation of the impact of function of problem behavior on effectiveness of the Behavior Education Program (BEP). *Education and Treatment of Children, 34*, 551-574.

Horner, R.H., & Sugai, G. (2000). School-wide behavior support: An emerging initiative (special issue). *Journal of Positive Behavioral Interventions, 2,* 231-232.

Horner, R. H., Sugai, G. M., & Anderson, C. M. (2010). Examining the evidence-based for school-wide positive behavior support. *Focus on Exceptional Children, 42,* 1–16.

Hosmer, D., & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). New York, NY: John Wiley & Sons, Inc.

Individuals with Disabilities Education Improvement Act. (2004). Public Law 108-446 (20 U.S.C. 1400 *et seq*.).

Ingram, K., Lewis-Palmer, T., & Sugai, G. (2005). Function-based intervention planning: Comparing the effectiveness of FBA function-based and non-function-based intervention plans. *Journal of Positive Behavior Interventions, 7,* 224-237.

Kazdin, A. E. (1995). Child, parent and family dysfunction as predictors of outcome in cognitive behavioral treatment of antisocial children. *Behavior Research and Therapy, 33,* 271-281.

Kraemer, H. C., Frank, E., & Kupfer, D. J. (2006). Moderators of treatment outcomes: Clinical, research, and policy importance. *JAMA, 296,* 1-4.

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry, 59*, 877-883.

Kratochwill, T. R., & Shernoff, E. S. (2004). Evidence-based practice: Promoting evidence-based interventions in school psychology. *School Psychology Review, 33*, 34-48.

La Greca, A.M. (1993). Social skills training with children: Where do we go from here? *Journal of Clinical Child Psychology, 22,* 288-298.

Landis, J. R., Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33:159-174.

Leedy, A., Bates, P., & Safran, S. (2004). Bridging the research-to-practice gap: Improving hallway behavior using positive behavior supports. *Behavioral Disorders, 29,* 130-139.

Lien-Thorne, S., & Kamps, D. (2005). Replication study of the First Step to Success early intervention program. *Behavioral Disorders, 31*, 18-32.

Loman, S. L., Rodriguez, B. J., & Horner, R. H. (2010). Sustainability of a targeted intervention package: First Step to Success in Oregon. *Journal of Emotional & Behavioral Disorders, 18*, 178-191. doi: 10.1177/1063426610362899

March, R. E., & Horner, R. H. (2002). Feasibility and contributions of functional behavioral assessment in schools. *Journal of Emotional and Behavioral Disorders, 10,* 158-170.

Mayer, G. R. (1995). Preventing antisocial behavior in schools. *Journal of Applied Behavior Analysis, 28,* 467-478.

McConaughy, S., & Ritter, D. (2008). Best practices in multimethod assessment of emotional and behavioral disorders. In A. Thomas & Grimes (Eds.), *Best practices in school psychology V* (pp.697-713). Betheseda, MD: National Association of School Psychologists.

McIntosh, K., Campbell, A., Carter, D. R., Dickey, C. R. (2009). Differential effects of a tier two behavior intervention based on function of problem behavior. *Journal of Positive Behavior Interventions, 11,* 82-93.

McLaren, E. M., & Nelson, C. M. (2009). Using functional behavior assessment to develop behavior interventions for students in Head Start. *Journal of Positive Behavior Interventions, 11,* 3-21.

McMahon, R. J., Wells, K. C., & Kotler, J. S. (2006). Conduct problems. In E. J. Mash & R. A. Barkley (Eds.), *Treatment of childhood disorders* (3rd ed., pp. 137-270). New York, NY: The Guilford Press.

Merrell, K. W. (2003). *Behavioral, social, and emotional assessment of children and adolescents.* Mahwah, NJ: Lawrence Erlbaum.

Merrell, K. W., Ervin, R. A., & Peacock, G. G. (2011). *School psychology for the 21st century: Foundations and practices* (2nd ed.) New York, NY: The Guilford Press.

Metzler, C., Biglan, A., Rusby, J., Sprague, J. (2001). Evaluation of a comprehensive behavior management program to improve school-wide positive behavior support. *Education and Treatment of Children, 24,* 448-479.

Miller, L. A., McIntire, S. A., & Lovler, R. L. (2011). *Foundations of psychological testing: A practical approach* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.

National Association of State Directors of Special Education. (2006). *Response to Intervention: Policy considerations and implementation*. Alexandria, VA: NASDSE, Inc.

Newcomer, L. L., & Lewis, T. J. (2004). Functional behavioral assessment: An investigation of assessment reliability and effectiveness of function-based interventions. *Journal of Emotional and Behavioral Disorders, 12,* 168-181.

Nunnally, J. C., & Kotsch, W. E. (1983). Studies of individual subjects: Logic and methods of analysis. *British Journal of Clinical Psychology, 22*, 83-93. doi: 10.1111/j.2044-8260.1983.tb00582.x

O'Connor, R., & Klingner, J. (2010). Poor responders in RTI. *Theory Into Practice, 49,* 297-304.

Overton, S., McKenzie, L., King, K., & Osborne, J. (2002). Replication of the First Step to Success model: A multiple-case study of implementation effectiveness. *Behavioral Disorders, 28*, 40-56.

Patterson, G., Reid, J., & Dishion, T. (1992). *A social learning approach: (Vol. 4). Antisocial boys*. Eugene, OR: Castalia.

Patterson, G. R. (1982). *A social learning approach: (Vol. 3). Coercive family process*. Eugene, OR: Castalia.

Payne, A. A. (2009). Do predictors of the implementation quality of school-based prevention programs differ by program type? *Prevention Science, 10,* 151-167.

Peng, C.-Y. J., So, T.-S. H., Stage, F. K., & St. John, E. P. (2002). The use and interpretation of logistic regression in higher education journals: 1988-1999. *Research in Higher Education, 43*, 259-293.

Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research, 96*, 3-14.

Petticrew, M., Tugwell, P., Kristjansson, E., Oliver, S., Ueffing, E., & Welch, V. (2012). Damned if you do, damned if you don't: Subgroup analysis and equity. *Journal of epidemiology and community health, 66*, 95-98.

Reid, J., Patterson, G., & Snyder, J. (2002). *Antisocial behavior in children and adolescents: A developmental analysis and model for intervention*. Washington, DC: American Psychological Association.

Rubin, A., & Babbie, E. (2001). *Research methods for social work*. Pacific Grove, CA: Wadsworth, Inc.

Scott, T. (2001). A school wide example of positive behavior support. *Journal of Positive Behavior Interventions, 3,* 88–94.

Serketich, W. J., & Dumas, J. E. (1996). The effectiveness of behavioral parent training to modify antisocial behavior in children: A meta-analysis. *Behavior Therapy, 27,* 171-186.

Sprague, J., & Perkins, K. (2009). Direct and collateral effects of the First Step to Success program. *Journal of Positive Behavior Interventions, 11*, 208-221. doi: 10.1177/1098300708330935

Strain, P., Young, C., & Horowitz, J. (1981). An examination of child and family demographic variables related to generalized behavior change during oppositional child training. *Behavior Modification, 5,* 15-25.

Streiner, D. L., & Norman, G. R. (2008). *Health measurement scales: A Practical guide to their development and use* (4th ed.). New York: Oxford University Press.

Sucuoglu, B., and Ozokçu, O. (2005). Kaynastirma ogrencilerinin sosyal becerilerinin degerlendirilmesi [Evaluation of social skills of students who were included in general classrooms]. *Ozel Egitim Dergisi 6*, 41–58.

Sugai, G., Horner, R., & Gresham, F. (2002). Behaviorally effective school environments. In Shinn, M. R., Walker, H. M., & Stoner, G. (Eds.) *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 315-350). Bethesda, MD: National Association of School Psychologists.

Sugai, G., Sprague, J., Horner, R., & Walker, H. (2000). Preventing school violence: The use of office discipline referrals to assess and monitor school-wide discipline interventions. *Journal of Emotional and Behavioral Disorders, 8,* 94–101.

Sumi, W. C., Woodbridge, M. W., Javitz, H. S., Thornton, S. P., Wagner, M., Rouspil, K., . . . Severson, H. H. (2012). Assessing the effectiveness of First Step to Success: Are short-term results the first step to long-term behavioral improvements? *Journal of Emotional and Behavioral Disorders.*Advance online publication. doi: 10.1177/1063426611429571

Sun, X., Briel, M., Walter, S. D., & Guyatt, G. H. (2010). Is a subgroup effect believable? Updating criteria to evaluated the credibility of subgroup analyses. *BMJ, 340,* 850-854.

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. K. S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities*, *34*, 33-58.

U.S. Department of Education. (2000). *Annual report to Congress on implementation of the Individuals with Disabilities Act*. Washington, DC: Office of Special Education and Rehabilitative Services.

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

Walker, H. M. (2004). Commentary: Use of evidence-based interventions in schools: Where we've been, where we are, and where we need to go. *School Psychology Review, 33*, 398-407.

Walker, B., Cheney, D., Stage, S., & Blum, C. (2005).Schoolwide screening and positive behavior support: Identifying and supporting students at risk of school failure. *Journal of Positive Behavior Interventions, 7,* 194-204.

Walker, H. M., Golly, A., McLane, J. Z., & Kimmich, M. (2005). The Oregon First Step to Success replication initiative: Statewide results of an evaluation of the program's impact. *Journal of Emotional & Behavioral Disorders, 13*, 163-172.

Walker, H. M., Golly, A., Severson, H. H., Stiller, B., Kavanagh, K., & Feil, E. G. (1997). *First Step to Success: Helping young children overcome antisocial behavior: Implementation guide*. Longmont, CO: Sopris West.

Walker, H. M., Kavanagh, K., Stiller, B., Golly, A., Severson, H. H., & Feil, E. G. (1998). First Step to Success an early intervention approach for preventing school antisocial behavior. *Journal of Emotional and Behavioral Disorders, 6*, 66-80.

Walker, H. M., & McConnell, S. R. (1988). *The Walker-McConnell Scale of Social Competence and School Adjustment: A social skills rating scale for teachers*. Austin, TX: Pro-Ed.

Walker, H. M., Ramsey, E., & Gresham, F. M. (2004). *Antisocial behavior in school: Evidence-based practices*. Belmont, CA: Wadsworth Publishing Company.

Walker, H. M., Seeley, J. R., Small, J., Severson, H. H., Graham, B. A., Feil, E. G., . . . Forness, S. R. (2009). A randomized controlled trial of the First Step to Success early intervention demonstration of program efficacy outcomes in diverse, urban school district. *Journal of Emotional and Behavioral Disorders, 17*, 197-212.

Walker, H. M., & Severson, H. H. (1992). *Systematic Screening for Behavior Disorders (SSBD): Technical manual*. Longmont, CO: Sopris West.

Walker, H. M., Severson, H. H., Todis, B. J., Block-Pedego, A. E., Williams, G. J., Haring, N. G., & Barckley, M. (1990). Systematic Screening for Behavior Disorders (SSBD) further validation, replication, and normative data. *Remedial and Special Education, 11*, 32-46.

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

Warren, J.S., Edmonson, H.M., Griggs, P., Lassen, S.R., McCart, A., Turnbull, A., et al. (2003). Urban applications of school-wide positive behavior support: Critical issues and lessons learned. *Journal of Positive Behavior Interventions, 5,* 80–91.

Ysseldyke, J., Burns, M., Dawson, P., Kelley, B., & Morrison, D. O. (2006). *School psychology: A blueprint for training and practice III*. Bethesda, MD: National Association of School Psychologists.

Zins, J.E., & Elias, M.J. (2006). Social and emotional learning. In G.G. Bear & K.M. Minke (Eds.), *Children's needs III: Development, prevention, and intervention* (pp. 1-13). Bethesda, MD: National Association of School Psychologists.