October 2020

Facilitating Early Intervention Through Teacher Training in Brief Functional Behavior Assessment

Casie L. Peet
University of South Florida

Facilitating Early Intervention Through Teacher Training in Brief Functional Behavior Assessment

by

Casie L. Peet

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in School Psychology
Department of Educational and Psychological Studies
College of Education
University of South Florida

Major Professor: Nathaniel von der Embse, Ph.D.
John Ferron, Ph.D.
Lise Fox, Ph.D.
Diana Ginns, Ph.D.

Date of Approval:
August 28th, 2020

Keywords: early childhood education, behavior problems, direct behavior rating, functional assessment

Acknowledgements

I will never be able to express my deepest gratitude to all the people who helped make this project possible. First, I want to thank my major professor, Nate von der Embse, for your steadfast support in both this project and in my development as a professional. Thank you for believing in me and pushing me outside my comfort zone. The lessons you taught me outside of academic content - about finding balance, accepting feedback with grace, and engaging in value driven action are lessons I do not take for granted. I want to thank Dr. Ferron for providing me with your calming guidance throughout this project, which was just as much about the importance of reflection and flexibility as it was about statistics. I am further grateful to Dr. Ginns and Dr. Fox for your feedback and insights that provided the rounded and critical considerations it takes to make research successful and meaningful. I want to give a special thank you to Mikayla Drymond, Sarahy Durango, and Ali Simons for joining my data collection team for this project. Your input and hard work helped make this project a reality and I couldn't have done it without you. Finally, thank you to my writing accountability partner and dear friend, Camille Hanks, for the countless coffee and writing dates and for being my hype woman throughout this project. I'd also like to thank the National Association of School Psychologists for their financial contributions to this project through the Graduate Student Research Grant.

As this project is a capstone, signifying the end of my doctorate, I want to take this moment to thank the people who have helped me reach this point. First and foremost, I want to thank my parents, John-David and Randi, whose love and unwavering confidence in me kept me grounded, focused, and sane. I could not have done this if you had not fostered my determination and taught me the value of hard work from a young age. Thank you to my brother and sister for

believing in me and keeping me humble. Thank you to my friends across the county who have supported me from a distance. Finally, I'm so incredibly grateful for the friendships graduate school has brought into my life, Camille Hanks, Mikayla Drymond, Nicholas Smith, Emily Wingate, Elizabeth Storey, Sarah Thoman, Brenna Joerger & David Rubio Jr. Thank you for being my cheerleaders, tutors, therapists, research consultants, mentors, and my family here in Tampa and for filling my moments in between writing and studying with laughter and joy.

Table of Contents

List of Tables

List of Figures

Abstract

Nearly 30% of students who enter state-funded early childhood education programs exhibit significant problem behavior, putting them at risk for long-term adverse behavioral and academic outcomes. Tier 2 behavioral interventions might not be effective for all when delivered in a one-size-fits-all fashion suggesting that individualizing behavioral intervention to the student's specific concerns may be. To do so, it is necessary to collect problem identification data indicative of each student's concerns and function of problem behaviors. This question is particularly pertinent in early childhood settings where educators have a wide range of training experiences and backgrounds. Early childhood teachers are essential partners in the consultative process; thus it is paramount they have the requisite skills in collecting accurate behavior data. This study sought to determine the effectiveness of several professional development protocols aimed at improving early childhood educator's foundational knowledge to increase their involvement in the consultative process.

Using a concurrent multiple baseline single-case design, this study evaluated the necessary level of training for early childhood educators to participate in the consultative process as data collectors. The researchers conducted the study with six preschool teacher-student dyads in the Southeastern United States. The baseline condition consisted of brief exposure to the data collection tool and took approximately 2-5 minutes. This was meant to represent the use of the tool in the absence of all training. The first training included a didactic on the basics of behavior and functional assessment. The second training consisted of a performance feedback component where teachers rate pre-recorded and pre-rated videos and then reviewed their assessment scores compared to the correct scores. Researchers conducted a systematic visual analysis, calculated

effect sizes using the Tau-U statistic, and ran multilevel models to determine the effectiveness of the two training protocols. All analyses were conducted for two variables: (1) Disruptive Behavior Agreement and (2) Consequence Agreement or the teacher's ability to determine the rate of disruptive behavior and the consequences or function of the behavior in an observation. Teachers showed high levels of agreement in baseline when rating disruptive behavior only. However, teachers showed high levels of disagreement when rating the functions or consequences of their students' behavior, which was largely unaffected by either training protocol. The frequency of disruptive behavior was statistically significant in every model as a covariate influencing agreement levels.

Results of this study suggest early childhood educators may have adequate foundational knowledge without additional professional development to serve within the consultative process as data collectors for frequency of disruptive behavior. The use of teachers as data collectors is critical as it can help build the consultative relationship and increase teacher buy-in for and engagement in the problem-solving process. However, more research is needed to determine the necessary levels of professional development for teachers to collect accurate and meaningful data on the function of behavior. School psychologists can use this information to engage early childhood educators with suitable professional development. Suggestions on future directions in research and implications of the tool in practice given the effect of behavioral frequency are discussed.

Chapter I: Introduction

Statement of the Problem

Students at risk for emotional and behavioral problems often have poor academic and social outcomes throughout school and into adulthood (Reinke et al., 2008; Sprague & Walker, 2000). Left untreated, emotional and behavioral problems intensify and manifest in poor outcomes into adolescence and adulthood, such as school failure, substance abuse, unemployment, violence, or suicide (Sprague & Walker, 2000). Early identification and intervention with these students are of increased importance, given these considerably adverse outcomes. Schools are the ideal setting to implement preventative population services given their broader and more frequent access to children than community services and other settings.

Schools have begun to respond to these adverse outcomes through the implementation of multi-tiered systems of support (MTSS), which is based on a foundation of problem-solving and data-based decision-making (Newton, Horner, Algozzine, Todd, & Algozzine, 2009). MTSS works from a preventative, public health model to implement interventions that are both preventative as well as reactive to problems that do arise. Generally, schools use a three-tiered approach that delineates universal curricula and practices for all students and secondary and tertiary supports for students who do not respond to the core academic, social-emotional, or behavioral curricula. Research has focused on the effectiveness of Tier 1 and Tier 3 interventions (Hawken, 2006). Substantially less has been done regarding Tier 2 and secondary supports. Consequently, schools struggle to support those students who do not respond to universal curricula but may not be severe enough to require intensive or individualized intervention.

This preventative approach to service delivery is particularly critical in early childhood settings, given the importance of early intervention. There has been a push for early childhood education and kindergarten readiness in recent years. Ackerman & Barnett (2005) describe the inherent issues in determining kindergarten readiness by age due to wide ranges in development. In most states, compulsory education begins at five or six. Developmental levels of five and six years old vary vastly between children and are heavily influenced by their environment and level of previous education (Ackerman & Barnett, 2005). School readiness is generally comprised of multiple domains including cognition and general knowledge, language, social and emotional skills, physical wellbeing, and motor development, as these skills have shown to be highly predictive of future academic success (Duncan et al., 2007; Hindman, Skibbe, Miller, & Zimmerman, 2010; Justice, Mashburn, Hamre, & Pianta, 2008; National Institute of Child Health & Human Development [NICHD] Early Child Care Research Network [ECCRN], 2005). Since preschool education enhances kindergarten readiness (Ackerman & Barnett, 2005), there has been increasing emphasis on early childhood practices and improving access to early childhood education programs such as Head Start. However, many educators in early childhood settings experience difficulty in delivering effective emotional and behavioral supports to prevent or improve problem behaviors (Fox & Little, 2001; McLeod et al., 2017). Nearly 30% of students who enter state-funded early education programs exhibit significant problem behavior (Barbarin, 2007) putting them at risk for long-term adverse outcomes including underachievement, grade retention, special education placement, and dropout and expulsion (Bulotsky-Shearer, Dominguez, & Bell, 2012; Reinke et al., 2008).

In alignment with public health models and multi-tiered systems of support, the Pyramid Model for Promoting Social-emotional Competence in Infants and Young Children (Fox, Dunlap, Hemmeter, Joseph, & Strain, 2003) was created to address concerns in early childhood

settings. Like other public health frameworks and MTSS, the Pyramid Model includes universal, secondary, and tertiary levels of support to address all students' needs, which will be described in greater detail in Chapter 2. At the universal and supplemental levels of support (Tier 1 and 2), schools often utilize a simplified one-size-fits-all approach to intervention selection (i.e., a standard behavior intervention when a student displays disruptive behavior). This simplified method of intervention selection helps schools quickly react to mild to moderate behavioral concerns with effective strategies for many students. Teachers often understand these interventions and have access to the materials, and therefore outside resources, such as student services staff, do not need to be involved in mild behavior problems. This eases the strain on the limited personnel and time resources in schools. However, many of these streamlined interventions might not be sufficient when delivered in a one-size-fits-all fashion (McIntosh, Campbell, Carter, & Rossetto Dickey, 2009), suggesting that individualizing behavioral interventions to the student's specific concerns may be needed (Campbell & Anderson, 2008). A synthesis of the current literature conducted by Dunlap et al. (2006) suggested that function-based interventions are more effective in early childhood for behavior problems than interventions that do not incorporate the behavior's function.

To tailor interventions to address a specific function (e.g. incorporating a break card to address an escape function), it is necessary to collect problem identification data indicative of each student's concerns and reinforcing consequences maintaining problem behaviors. Unfortunately, many existing problem identification assessment tools (e.g., systematic direct observations) lack the feasibility required for use at Tier 2 due to the resources necessary to conduct them. For example, systematic direct observations (SDO) is a method of observation that uses interval time sampling to calculate rates of behavior and consequences. However, this method requires personnel trained in SDO to conduct multiple observations in the classroom.

Conversely, indirect functional assessment tools have been found to yield inconsistent results (e.g., Iwata, DeLeon, & Roscoe, 2013; Zarcone, Rodgers, Iwata, Rourke, & Dorsey, 1991). This lack of adequate tools is magnified in early childhood settings, as less work has been done in early childhood assessments, specifically regarding function. The most common assessment tools used in early childhood aim to detect the presence of clinical levels of behavior problems, rather than determine the function of the behavior or information helpful for intervention planning. To date, no assessment tools aimed at identifying the function of the problem behavior have been validated for use in early childhood.

The *Intervention Selection Profile-Function* (ISP-Function), a novel functional behavior assessment tool, has demonstrated preliminary evidence (Kilgus, Kazmerski, Taylor, & von der Embse, 2016) as an efficient method for teachers to assess a student's behavior and determine a probable function through direct observations of the student. The tool is currently under ongoing validation within an elementary population in a 4-year project funded by the Institute of Educational Sciences. This tool has yet to be examined in a preschool population, where early identification and intervention efforts can be maximized through the youngest students in our education system. Therefore, to improve early identification and prevention of future social and academic concerns, this project evaluated this tool's utility in preschool classrooms and the degree of teacher training necessary for a teacher to reach accuracy. By streamlining the process of implementing interventions on a continuum, preschools can move towards a more preventative service delivery model. Effective use of a brief problem identification tool before Tier 3 in preschools can help promote prevention and early identification. Further, by utilizing this process, preschools can support efforts towards kindergarten readiness and improve students' overall likelihood of success.

Purpose of the Study

The present study evaluates the utility of the ISP-Function in preschool settings and, more specifically, the level of training necessary for teachers to use the tool with psychometric adequacy. The ISP-Function is intended to help individualize Tier 2 behavioral interventions and offer appropriate problem identification information on student behavior problems for early educators. Early identification is particularly important at the preschool level, given the poor outcomes of students who show behavior risk at such an early age. Further, without teacher training, poor data quality may lead to inaccurate or incorrect intervention decisions. The primary aims of this study are 1) to determine the extent to which rater training is necessary to support the accuracy of ISP-Function data collection and use in preschools, and 2) to identify which of several training protocols is the most feasible and practical approach to rater training. The results of this study seek to support educators in understanding behavioral principles to assist in accurate behavioral data collection and, ultimately, intervention selection and implementation.

Utilization in Preschools

School psychologists depend on psychometrically sound assessments to make decisions on intervention selection, continuation, and termination as well as high stakes decisions such as placement or special education qualification. Currently, there are no widely used and evidence-based Tier 2 assessment tools to inform behavioral intervention selection resulting in little to no individualization for Tier 2 practices. Implementing this tool in preschools and early childhood programs can help school psychologists and student services teams make intelligent and thoughtful decisions based on the student's concerns and function of their behavior as we work within the problem-solving framework.

As aforementioned, effective implementation, psychometric adequacy, and appropriate decisions made with this tool may be influenced by rater training. Successful implementation, including educator training of the ISP-Function, can strengthen Tier 2 practices by allowing staff to individualize Tier 2 interventions and based on the function of the behavior. Further, preschools can use this tool to evaluate trends across students to inform class-wide or school-wide behavior practices. This can facilitate a shift from responsive practices at Tier 3 and towards preventative strategies at Tier 1 and 2 levels. Prevention is particularly crucial in the early childhood sector as preschools are ideal environments for early intervention, preventative practices, and improving kindergarten readiness in students.

In the future, this tool may be used widely by programs striving to build out their MTSS and enhance their problem-solving procedures. Because data and data-based decision making is a fundamental building block of response to intervention (RtI) and problem-solving, this tool can assist preschools in addressing the lack of data used in between universal screening (Tier 1) and comprehensive functional behavioral assessments or evaluations (Tier 3). This research may inform how to best implement this tool with fidelity, including necessary levels of teacher training. Future research can expand on the utility and social acceptability of this tool, as well as the appropriateness of decisions made based on the data. This tool may improve MTSS efforts (e.g. the Pyramid Model) and help early childhood programs address challenging behaviors at varying intensity levels.

Specifically, in preschools, this tool may be integrated within the Pyramid Model as a complementary strategy to use with current universal and supportive instructional practices. Within the Pyramid Model, preschool teachers are encouraged to provide social-emotional learning to all students. Students who display behavior problems are provided a higher intensity of social-emotional instruction by increasing the frequency, dose, or precision of instruction.

This provides students who may have a social-emotional or behavioral skill deficit but does little to address motivational deficits that may be more appropriately met with contingency management interventions. At the tier 2 level, the function of the behavior is not addressed within the Pyramid Model framework, and there is no current explicit guidance related to using functional assessment to design instruction. While the Pyramid Model does incorporate functionally informed interventions at Tier 3, as described in more detail later, the process requires comprehensive assessment methods that make it difficult to implement at a tier 2 level. Addressing the function of behavior in conjunction with the social-emotional learning guidance outlined in the Pyramid Model, may help to improve student behavior that is unresponsive to solely instructional interventions prior to moving to the resource intensive tier 3 process. By diversifying available intervention types at a tier 2 level prior to a more individualized and resource-intensive process, such as an evaluation, preschools may be able to improve early intervention practices, and thus preventing more severe behavior that necessitates an intensive support plan.

Research Questions

This assessment tool is unique, given its efficiency and ability to be completed by classroom staff rather than having outside personnel (e.g., school psychologist) conduct the assessment. However, to utilize preschool teachers in collecting accurate and adequate data, training in the use of the tool and foundational knowledge may be necessary. Preschool teachers have vastly different backgrounds and training, which might present issues in the quality of data collected. In the literature review that follows, evidence will be reviewed, suggesting preschool teachers require additional training in functional behavior assessment methods (Martin, 2016). Thus, the first research question seeks to answer if there is a functional relationship between teacher training and the accuracy of ISP-Function data collected by preschool teachers. The

researcher hypothesized that there would be a functional relationship between teacher training and the accuracy of ISP-Function data and that the implementation of teacher training will increase the accuracy of ISP-Function scores.

The scarcity of resources in schools calls for the following two aims of this project. While more training is expected to improve accuracy, the second research question aims to determine if one of two training components is more effective in reaching ISP-Function accuracy. Two training protocols were tested: a) Behavior Basics and b) Performance Feedback training. Previous studies have found improvements in accuracy following performance feedback (Kilgus et al., 2017). The researcher hypothesizes that Performance Feedback training would be the most effective method of preschool teacher training on ISP-Function scores. Finally, if both training components are necessary for the teacher to reach accuracy using the ISP-Function, the final research question evaluated whether there is a difference in the order of the two training components. It was expected that there would be an effect on ISP-Function scores based on the order of the trainings implemented and that the Behavior Basics training protocol followed by the Performance Feedback protocol would yield the highest degree of ISP-Function accuracy. This hypothesis was based on the findings from Martin (2016), indicating that early childhood teachers may need foundational knowledge regarding functional behavior assessments. Therefore, it was expected that providing the Behavior Basics protocol first may improve accuracy overall.

In summary, this project aimed to answer three interrelated research questions related to the utility and implementation of the ISP-Function in preschool settings:

1. To what extent does teacher training improve the accuracy of ISP-Function data collected?

2. Which training component is the most effective method of training on the accuracy of ISP-Function data collected?
3. Is there a difference between the order in which the training components are implemented on the accuracy of ISP-Function data collected?

Summary

Behavior problems continue to be a problem in our schools. Early childhood settings are of increased importance, given the value of early identification and intervention. The push for school and kindergarten readiness has highlighted the need for quality early childhood education to address overall readiness that includes academic, social-emotional, and behavioral readiness. Further research is necessary to address how we can utilize sound assessments within a preventative, multi-tiered framework to address students' behavioral needs. The ISP-Function is a promising new tool that utilizes direct behavior rating methodology to assess disruptive behavior and consequences to cater Tier 2 interventions to student function. The novelty of this tool using classroom staff as raters raises the need to evaluate the degree of teacher training necessary to collect accurate data in preschools. This question is particularly pertinent in early childhood settings where educators have a wide range of training experiences and backgrounds. Thus, this research seeks to answer the above research questions regarding the practical implementation of this tool in early childhood settings.

Chapter II: Review of the Literature

Behavior Problems

The mental health epidemic in America is of increasing concern for public health fields and education. Approximately one in five students in America meet criteria for a diagnosable mental disorder. In a landmark study of the prevalence of mental health (Merikangas et al., 2010). Merikangas et al. (2010) found that the overall incidence of disorders with severe impairment was 22.2% in a large nationally representative sample of adolescents aged 13-18 in the United States. The most common conditions were anxiety disorders (31.9%), behavior disorders (19.1%), mood disorders (14.3%), and substance use disorders (11.4%; Merikangas et al., 2010). Finally, the median age of onset in this sample ranged from 6-15 years old for various disorder classes, and the probability of being diagnosed increases with age, specifically in adolescence. Despite these dismal findings, research has found that early identification and intervention may help reduce the number of mental health concerns in older children and adolescents (Kauffman & Hallahan, 2011).

Developmental Cascades

For years, research has sought to identify predictors of problematic behaviors, mental health problems, and poor adjustment to enhance early identification efforts. Research efforts to assess the relationship between potential childhood features that predict adult adjustment resulted in the development of the developmental cascades model. Developmental cascades are a phenomenon in which early difficulties in one ecological domain may have far-reaching and significant effects in another developmental area later (Garmezy, Masten, & Tellegen, 1984; Masten, Desjardins, McCormick, Kuo, & Long, 2010; Masten et al., 2005; Obradović, Burt, &

Masten, 2009). Specifically, externalizing problems early in childhood negatively influenced academic success by adolescence, which consequently increased internalizing symptoms in early adulthood (Masten et al., 2005). Students at risk for emotional and behavioral disorders have lower grades, test scores, and higher school dropout rates (Reinke, Herman, Petras, & Ialongo, 2008; Sprague & Walker, 2000). Further, Christner, Forrest, Morley, & Weinstein (2007) found a cyclical effect in that emotional and behavioral issues can influence learning, and academic difficulties can then exacerbate emotional and behavioral problems (Christner et al., 2007). Emotional and behavioral disorders are associated with academic failure and many adverse societal outcomes (Lehr & McComas, 2005; Masten et al., 2005). Additional research has associated untreated emotional and behavioral problems with substance abuse, associating with peers involved in risky behaviors, teen pregnancy, chronic mental health problems, employment difficulties, violence, and suicide (Sprague & Walker, 2000).

The evidence for the significant influence of externalizing problems on future academic and life success is an impetus for a shift toward actively addressing social-emotional and behavioral needs in schools. The mission of schools is to provide students with the tools to lead successful lives and further, that success is a product of social and emotional competence and academic achievement (Doll & Cummings, 2008). As such, more comprehensive definitions of school success have begun to include both academic and social-emotional domains (Roeser, Eccles, & Sameroff, 1998). These definitions are supported by previous research that has linked the reciprocal and inseparable nature of academic success and mental health (Suldo, 2016). Given the increasing emphasis on social-emotional wellbeing and behavioral competence, schools are beginning to address and prevent behavior problems through preventative frameworks and practices. Universally, these practices Positive Intervention Behavior Supports (PBIS), Universal Social Emotional Learning (SEL) curricula, and using universal screening to

identify students who need more support. At more individualized levels such as Tier 2 and 3, schools use evidence-based social-emotional and behavioral interventions to address students' wellbeing in addition to just their academic success.

This project specifically focuses on the collection of disruptive behavior data in preschool-age students. Disruptive behavior, for this project, is broadly defined as behaviors that interfere in the learning of the target student or other students in the classroom. Specific examples and nonexamples are often determined by the teacher and normative rules of the classroom. For example, some classes allow students to move around to get water or go to the bathroom without asking, while in other classrooms, this would be considered disruptive, and students must obtain permission before leaving a designated area. However, there are universal examples that typically occur in the preschool population such as calling out, excessive movement, arguing, tantrums, or other conduct that disrupts the continuity of the learning environment (Dominguez, Vitiello, Fuccillo, Greenfield, & Bulotsky-Shearer, 2011).

Treatment of Behavior Problems

A recent meta-analysis of single-case design studies conducted by Walker, Chung, & Bonnet (2018) evaluated the effect of function-based interventions on student behavior problems. This study found that overall, function-based interventions resulted in a decrease of challenging behavior and increased appropriate behavior (Walker et al., 2018). This is consistent with previous meta-analyses that support the use of function-based approaches for behavior problems in schools (e.g., Gage, Lewis, & Stichter, 2012). Through a behavioral lens, repeated behavior occurs for a specific reason or purpose, known as the function (Cooper, Heron, & Heward, 2007). The function of a behavior, or reason that the behavior occurs, is critical to consider and understand when assessing and treating problematic behavior. Research has long substantiated the link between identifying a function of behavior as an essential component to facilitate

effective behavioral interventions (Borgmeier, Loman, & Strickland-Cohen, 2017; McIntosh, Brown, & Borgmeier, 2008; O'Neill et al., 2015).

Typically to help guide the development and implementation of functional-based interventions, an evaluation process known as a Functional Behavioral Assessment (FBA) is conducted to help the student intervention team decide what intervention is most appropriate for the student. A FBA is a comprehensive and time-intensive assessment process typically consisting of both direct (i.e., observations) and indirect (i.e., rating scales and teacher/parent interviews) methods. An FBA aims to determine all environmental factors that evoke and maintain problem behavior (Anderson, Rodriguez, & Campbell, 2015). This includes both antecedent situations that predict the problem behavior and consequences that follow the problem behavior (the function).

Antecedents may be events (e.g., being redirected) or setting characteristics (e.g., when the classroom is loud). Both of these types of antecedents can help teachers, clinicians, or parents predict when behavior is more likely to occur and control settings, give prompts, and prevent problem behavior from occurring. FBA's also aim to determine the maintenance consequence of problem behavior. For example, if a child consistently receives adult attention following problematic behavior, a reasonable hypothesis is that it is displayed to gain that adult attention. Interventionists can use this information to develop an intervention that removes the reinforcing consequence (e.g., adult attention) following problematic behavior and provides it following desirable behavior.

To gather this information, a comprehensive FBA includes multiple methods of assessment and various informants providing information. As such, many different tools are utilized as part of an FBA. Anderson et al. (2015) conducted a recent review of the current state of FBA's in schools. First, FBA's were used most often with students with intellectual

disabilities and autism spectrum disorder (ASD) and most often to address physical aggression, self-injurious behavior, and disruption of property (Anderson et al., 2015). This is a logical finding as schools are more likely to use time and resource-intensive assessment methods such as an FBA for severe behaviors (e.g., physical aggression) rather than off-task behavior or non-compliant behavior.

A majority of publications in this review (56.7%) used more than one type of FBA, including direct and indirect measures (Anderson et al., 2015). The most frequently used type of indirect assessment was a functional assessment interview form followed by rating scales (i.e., Motivation Assessment Scale), and problem behavior questionnaires. Direct observations were also frequently used (Anderson et al., 2015). Importantly, this review examined the FBA methods used for research done in schools, which may differ from actual practice occurring in schools. However, the above results mirror the practices outlined in the Pyramid Model (Fox, Dunlap, Hemmeter, Joseph, & Strain, 2003) for addressing chronic behavior problems in early childhood settings for developing individualized support plans at tier 3. The Pyramid model describes the functional assessment process to include reviews of Behavior Incident Report (BIRS) data and patterns, interviews with the teacher and parent, as well as direct observations. A summary of these data is combined and discussed with the team to discuss intervention development and implementation. Prevent-Teach-Reinforce for Young Children (PTR-YC; Dunlap, Wilson, Strain, & Lee, 2013) outlines a functional assessment process that includes establishing goals and objectives, a rating scale to collect baseline and progress monitoring data, and a functional checklist to understand antecedents, function and replacement behaviors. However, these procedures and strategies to incorporate function to individualize interventions are reserved for students with the most significant behavior problems at tier 3. At the tier 2 level,

social emotional learning instruction is provided at a higher frequency or intensity. There is no individualization that occurs at the tier 2 level as currently described in the Pyramid Model.

Research has found medium-sized effects in function-based interventions that have been informed by FBA's (Gage et al., 2012; Trussell, Lewis, & Raynor, 2016). Hawken, O'Neill, & MacLeod, (2011) evaluated the effect of using function to tailor the Behavior Education Program (BEP), a standard Tier 2 intervention in elementary and preschools. This study indicated that BEP is more effective for students with peer versus adult attention maintained behavior problems (Hawken et al., 2011). Additionally, BEP was effective for students whose behavior functioned to access tangibles or preferred activities, and results were mixed for the students with escape-maintained behavior. In contrast, Bruni et al. (2017) found a small but insignificant difference between function-based informed interventions and interventions not informed by the function. Specifically, in early childhood, Dunlap et al., (2006) synthesized the literature regarding the presence, prevention, and treatment of behavior problems in early childhood. An aggregation of results from descriptive, quasi-experimental, and experimental peer-reviewed studies using single-subject designs suggested that function-based interventions effectively reduce behavior problems in young children (Dunlap et al., 2006). Another study found that interventions informed by functional analyses decreased problem behavior in preschool and HeadStart classrooms (Dufrene, Doggett, Henington, & Watson, 2007). Overall, early childhood research supports that function-based interventions are both more effective and robust than interventions that are not informed by a functional assessment (Dunlap & Fox, 2011; Dunlap et al., 2006).

However, while functional assessments have support in the literature for addressing behavior problems, it may not be feasible for schools to conduct such a time and resource-intensive evaluation on all students with behavioral concerns. FBA's are typically reserved for students at Tier 3 (~5% of students), although the MTSS framework suggests 15-20% of students

will struggle behaviorally and need additional support. In schools, specialized personnel are required to come into the classroom to conduct observations, speak with parents and teachers, obtain rating scales, and compile the results. Following, a team typically meets to discuss the results, plan of action, and create a behavior intervention plan. This process is defined within the Pyramid Model for Early Childhood settings and similarly in Prevent-Teach-Reinforce for Young Children (PTR-YC). Indeed, this complete process may be too intensive in terms of time and personnel required to conduct full FBA's outside of the most severe (Tier 3) cases in a school. Despite these promising findings, interventions in isolation rarely maintain their effectiveness if there is no robust system in place (Kauffman & Hallahan, 2011). It is critical for the whole system to have an active process and set of procedures for preventing, identifying, and addressing behavioral challenges in schools. Having a robust universal system can also prevent the need for more intensive interventions (Doll & Cummings, 2008), which is particularly crucial given scarce resources.

Early Childhood Education

As aforementioned, school readiness definition generally includes multiple domains including cognition and general knowledge, language, social and emotional skills, physical wellbeing, and motor development as these skills are highly predictive of future academic success (Duncan et al., 2007; Hindman, Skibbe, Miller, & Zimmerman, 2010; Justice, Mashburn, Hamre, & Pianta, 2008; National Institute of Child Health & Human Development [NICHD] Early Child Care Research Network [ECCRN], 2003). This comprehensive definition of school readiness and attention to social-emotional and behavioral factors, in addition to cognitive ability and general knowledge, underscores the importance of supporting student behavior from a preventative and proactive standpoint. One study found that behavior regulation in preschool children was significantly predictive of emergent literacy, vocabulary, and math skills, as well as

growth in these areas across the prekindergarten year after controlling for demographic variables (McClelland et al., 2007). Readiness is primarily influenced by environmental factors, including home and community factors, in addition to previous academic, social, emotional, and behavioral experiences (Ackerman & Barnett, 2005; Regenstein, Connors, Romero-Jurado, & Weiner, 2017). Children enter kindergarten with inequitable levels of prekindergarten educational experiences (Barnett & Yarosz, 2007), and consequently varying numeracy, literacy, and language skills (Zill, 1999). Since kindergarten readiness can be enhanced by preschool education (Ackerman & Barnett, 2005), there has been increasing emphasis on practices in early childhood and improving access to early childhood education programs such as Head Start as these are vital settings to target students for early identification and intervention.

Because early childhood education is not compulsory, the types and availability of early childhood settings range widely. Traditionally, early childhood education was provided privately and outside of school districts or state-funded programs. In the push towards universal school readiness, Head Start Programs were developed as an anti-poverty program to help address achievement gaps by providing low-income families access to preschool, who typically could not afford private preschool education. Head Start and Early Head Start programs are funded federally and are available in every state in the US. Early Head Start programs are available for children under three. Head Start is open for children 3-5 years old. While Head Start is only available to low-income families, some states are beginning to offer voluntary preschool through state funds (e.g., State Department of Education) for students regardless of income status. For example, in Florida, Voluntary Prekindergarten Education Program (VPK) is available at many schools for four-year-olds. These types of programs vary in their accessibility, availability, and qualifications by state.

Within early childhood settings, the qualifications necessary for teachers vary as well (Garver, 2020). Private preschools are not required by law to require any specific degree, so teacher qualifications vary the most in private settings. Head Start programs are federally funded, and there are minimum degree requirements for teachers in Head Start programs. According to the Head Start Program Fact Sheet (2017), 96% of Head Start teachers had at least a two-year degree in Early Childhood Education (ECE) or related field. However, 73% have a bachelor's degree or higher in ECE or a related field (Head Start Program Facts Fiscal Year, 2017). Qualifications for state-funded preschools also vary by state. In Florida, there are several different credentialing routes to becoming a VPK provider that includes having an Associate's Degree or Bachelor's Degree in ECE or taking a series of trainings through the Florida Department of Education. For example, one such route is obtaining the Child Development Associate (CDA) Credential, which includes providing evidence of 600 hours of work experience with the age group, a professional portfolio, an exam, and a verification visit. Given this range of options, teacher knowledge, expertise, training, and experience vary vastly between and within early childhood programs. This presents a particular challenge when faced with students who struggle behaviorally in the classroom. The range of training requirements does not guarantee that early childhood educators are equipped with the skills to handle the significant disruptions that are often seen in classrooms. To support teachers in working with these students, schools as a whole typically work within a multi-tiered framework to appropriately allocate student service resources based on student needs.

Multi-Tiered Systems of Support

Many school systems have started to focus on prevention-oriented service delivery in the form of multi-tiered systems of support (MTSS) in recognition of the importance of early identification and intervention (The National Research Council and Institute of Medicine, 2009).

MTSS is a framework that helps schools to allocate necessary resources to both prevent and intervene with students at risk for behavior, academic, and social-emotional problems. MTSS is founded upon problem-solving logic and data-based decision making (Newton, Horner, Algozzine, Todd, & Algozzine, 2009). Generally, schools use a three-tiered framework to address all student needs: Tier 1 (universal) Tier 2 (supplemental) and Tier 3 (individualized).

Regarding intervention, Tier 1 intervention refers to the core curriculum or universal supports put in place for all students. For students who do not respond to Tier 1 supports or who show higher levels of risk for problems or school failure, Tier 2 supports are added (e.g., small group, differentiated instruction). Finally, for students who failed to respond adequately to Tier 2 supports or display significant levels of risk, individualized supports are often put in place in the form of an individualized education plan (IEP). Similar models are employed within early childhood settings. One model that has been well researched is the Pyramid Model for Promoting Social-Emotional Competence in Infants and Young Children (Fox, Dunlap, Hemmeter, Joseph, & Strain, 2003). The Pyramid Model delineates a similar three-tiered structure for students in early childhood settings and helps preschools promote social-emotional and behavioral wellness. The following sections describe best practices for preventing and addressing problem behavior in schools through a tiered framework, and specifically within the Pyramid Model.

Universal Prevention. MTSS typically establishes three primary levels of supports: universal, supportive, and tertiary supports. Universal supports are often referred to as universal prevention or core curriculum. School-Wide Positive Behavior Intervention Supports (SWPBIS; Sugai & Horner, 2002) is the gold standard in addressing behavioral and externalizing issues at the universal level. SWPBIS is a framework of tiered supports based on research in applied behavior analysis that emphasizes the prevention of behavior problems through explicit teaching of appropriate behavior and positive reinforcement. More than 20,000 schools in America use

SWPBIS (Horner et al., 2014). The core features of SWPBIS are to establish school-wide expectations for behavior, provide frequent positive reinforcement for positive behaviors and consistent responses to negative behavior. A final and imperative feature of SWPBIS is the use of data to monitor the school's progress as a whole as well as individual students to inform the type and intensity of supports provided (Merikangas et al., 2010). A vast amount of literature has supported the effectiveness of SWPBIS (e.g., (Freeman et al., 2016; Horner et al., 2014; Sugai & Horner, 2002). Positive student outcomes commonly associated with the implementation of SWPBIS include reduced office discipline referrals (Bradshaw, Mitchell, & Leaf, 2010; Kelm, McIntosh, & Cooley, 2014), reduced suspensions (Barrett, Bradshaw, & Lewis-Palmer, 2008; Bradshaw et al., 2010), increased attendance (Freeman et al., 2016; Pas & Bradshaw, 2012), increased academic achievement (Pas & Bradshaw, 2012; Simonsen et al., 2012), and improvements in social-emotional competence (Bradshaw, Waasdorp, & Leaf, 2012; Cook et al., 2015). Other positive outcomes include enhanced teacher efficacy perceptions (Kelm & McIntosh, 2012; Reinke, Herman, & Stormont, 2013), and improvements in school climate (Bradshaw, Koth, Thornton, & Leaf, 2009; Caldarella, Shatzer, Gray, Young, & Young, 2011).

Similarly, the Pyramid Model describes Universal Promotion Practices to create nurturing and responsive relationships and a high-quality supportive environment. The foundation of these promotion practices is the creation of strong relationships with children and families and teachers. Critical components are described to help schools establish the necessary groundwork for these practices. The model emphasizes the importance of specific teaching practices that are associated with positive outcomes such as joining in play with students, providing descriptive praise for appropriate behavior (e.g., “thank you for sitting quietly”), and developing relationships with families (Cox, 2005; National Research Council, 2001). The second aspect of universal practices ensures a high quality and supportive environment by

underscoring the importance of utilizing developmentally appropriate structures, including scheduling, transition, explicit instruction of expectations, and providing engaging activities that incorporate active participation and age-appropriate activities (e.g., group sing-a-longs rather than worksheets).

Targeted Supports. Schools and early childhood programs vary widely on practices used to support students who display behavioral risk and how these students are identified. Based on the MTSS framework, Tier 2 interventions should ideally provide supplemental targeted supports based on the presenting concern to prevent the problem from needing individualized (Tier 3) services and ameliorating negative outcomes for students who show low-intensity problem behavior (Hawken, Adolphson, Macleod, & Schumann, 2009). Schools typically aim for Tier 2 interventions to be efficient, general, integrate seamlessly within school infrastructures, increasing the availability of the services to students (Hawken et al., 2009). While research has lagged in the area of Tier 2 interventions as compared to Tier 1 and Tier 3 interventions (Lane, Oakes, Ennis, & Hirsch, 2014), current literature has separated Tier 2 interventions into two distinct categories: instructional interventions and contingency management interventions.

Instructional interventions are interventions in which students are taught positive skills (e.g., cooperation, organization). These are appropriate for students who engage in problem behavior because they lack the skills required to engage in positive behavior that would replace the undesirable behavior (Gresham, Elliott, Vance, & Cook, 2011). For example, a student may frequently argue with peers during group work because he or she does not have adequate cooperation skills. Thus, an appropriate intervention would be to teach him or her basic cooperation skills to replace the negative, argumentative behavior. This instruction often occurs in a small group ‘social skills group' format conducted either by the teacher or student services

personnel (e.g., school counselor). Research has supported the effectiveness of Tier 2 skills interventions to target social skills (DiPerna, 2006; Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Gresham, Cook, Crews, & Kern, 2004). Many well-established curricula address a wide range of skills such as The Incredible Years (Webster-Stratton et al., 2008) and Second Step (Committee for Children, 2011).

However, sometimes students demonstrate a "performance deficit," in that they have the appropriate skills necessary to engage in positive behavior but lack sufficient motivation or display the appropriate behavior with insufficient frequency (Gresham, Sugai, & Horner, 2001). In this case, contingency management interventions are considered more appropriate. Contingency management interventions manipulate the consequences that follow a student's behavior (Gresham, 2011). This strategy is founded in applied behavior analysis (ABA; Cooper, Heron & Heward, 2007). According to applied behavior analysis theory, problem behavior is often a result of receiving rewarding consequences (e.g., attention) more frequently for problem behavior (e.g., calling out in class) than their positive behavior (e.g., raising hand). As such, the goal of a contingency management intervention is to restructure the consequences that follow both positive behaviors and undesirable behaviors. This restructuring must remove not only rewarding consequences that follow undesirable behavior but also provide rewarding consequences following the positive replacement behavior to gradually reshape the student's behavior. These types of interventions lead to decreases in ODR rates (McIntosh et al., 2009), increases in academic engagement (Hawken & Horner, 2003), and reductions in problem behavior (Campbell & Anderson, 2008).

However, there is a significant gap in the literature regarding Tier 2 contingency management intervention in early childhood. Instead, at Tier 2, early childhood settings more often address social-emotional skills building. The Pyramid Model indicates social-emotional

instructional supports for students who appear to be at risk for challenging behaviors. These social-emotional skill instructions generally focus on identifying and expressing emotion, social problem solving, self-regulation, social interactions, handling frustration and anger, and friendship skills. Classroom teachers provide this instruction. The Pyramid Model typically uses these strategies to prevent persistent problem behavior from occurring in students who show occasional problem behavior or signs of a social-emotional deficit. Students who display persistent problem behavior within early childhood settings, and are not responsive to skill instruction, are provided with intensive and individualized support based on a functional assessment.

An alternative Tier 2 intervention is the Behavioral, Emotional, and Social Training: Competent Learners Achieving School Success (BEST in CLASS; Sutherland, Conroy, Abrams, & Vo, 2010). The intervention focuses on arming teachers with evidence-based instructional strategies. Specifically, the program focuses on (1) establishing rules, expectations, and routines, (2) behavior-specific praise, (3) pre-correction and active supervision, (4) opportunities to respond and instructional pacing, (5) teacher feedback, and (6) home-school communication (Vo, Sutherland, & Conroy, 2012). While this intervention has shown promising effects in a randomized control trial (ES ranging from 0.44-0.46; Sutherland, Conroy, Algina, Ladwig, Jessee, & Gyure, 2016), this intervention also lacks consideration of the function of the behavior. Participants in this study were selected based on teacher report, the identification of emotional and behavioral disorder risk through a teacher-completed screening tool, and average developmental scores. No individualized data was collected to cater the intervention for specific students in the design of the intervention across studies. This is common for most Tier 2 interventions in early childhood and throughout elementary grades as well.

Finally, First Step Next (an updated version of First Steps for Success; Hill et al., 2018) is a research-based intervention program to address disruptive behavior in early to late childhood. The core components of the intervention include direct instruction, group and individual contingency management, peer and home collaboration with rewards, and an emphasis on labeled praise. Previously, the First Steps for Success program included a component called "homeBase" in which there was a function-based component and increased parental involvement. This portion was removed in the First Steps Next update due to lack of fidelity of implementation. There is also an optional Functional Behavior Assessment as part of the screening process. However, the elimination of the homeBase program and optional status of the Functional assessment procedures, illustrate the lack of attention to function in many early childhood intervention programs. The one-size-fits-all approach for contingency management is intended to be less resource intensive in both assessment and individualization of the intervention. Comparable social-emotional skills training and teacher coaching are excellent initial steps for addressing problem behavior. However, students who do not respond may require a more individualized approach that incorporates the function of behavior. Traditionally, all individualization occurs at Tier 3.

Intensive and Individualized Supports. When students fail to respond with adequate progress to targeted supplemental supports or show severe behavior problems that necessitate individualized attention, Tier 3 supports are then developed and put in place. Several interventions can be used depending on the referral concern, intensity, and other mediating factors influencing the student. A common form of individualized intervention to address behavior problems is a behavior intervention plan (BIP), which is a series of prevention, instructional, and reinforcement strategies provided for the student by the teacher or student services team member. Tier 3 supports are highly individualized based on the student's unique

presentation, contributing factors (e.g., lack of sleep), the function of the behavior (e.g., attention-seeking), potential skill deficits (e.g., lack of social skills) and may also be developed through an ecological perspective to consider familial, cultural, or societal considerations. The Pyramid Model uses a team to conduct a functional assessment (e.g., observations and interviews with teachers and parents) that informs an individualized behavior support plan. This support plan incorporates the function of the behavior using consequence and prevention strategies and teaches replacement strategies. A strength of this model is that it also includes considerations of broader ecological factors of the child's context, including home and family life that may influence the student's development or progress. The data must continue to be collected throughout Tier 3 intervention to ensure progress continues and plans are revised as necessary.

Importance of Assessment within Preventative Frameworks

As aforementioned, MTSS, and specifically the Pyramid Model for Early Childhood settings, is founded upon data-based decision methodology and therefore is contingent upon consistent and quality assessment procedures. To determine which students require which level of tiered support, assessment practices are utilized throughout the problem-solving process at each tier. Assessment practices for academic needs are far more advanced than they are for social-emotional and behavioral needs. To illustrate, academic assessments are used at Tier 1 through standardized testing, mandatory quarterly evaluation, and unit classroom tests. All students complete these assessments, and teachers can use this data to determine which students are struggling to meet benchmarks. At Tier 2, students who have been identified as struggling may receive additional group instruction, brief assessments (e.g., curriculum-based measures) to assess growth more frequently than other students to ensure that targeted and individualized interventions are effective and closing the gap between struggling students and their peers. At Tier 3, an evaluation is often conducted to determine if there is a need for special education

services, which consists of IQ testing, achievement testing, and a series of observations, behavior rating scales, and adaptive assessments. For social-emotional and behavioral needs at Tier 1, universal screening practices, or existing data (e.g., behavior incident reports) are often used to identify students at risk for emotional and behavioral issues. At Tier 3, observations, interviews, or rating scales completed by parents or teachers are given to determine a more specific problem and may also be used to progress monitor any interventions. For severe behavioral problems, a functional behavioral assessment may also be conducted. However, no standardized assessment methods are being used at Tier 2 to address social-emotional and behavioral needs.

Universal Screening. Many schools use existing data such as behavior incident reports system (BIRS), attendance data, records of suspensions and expulsions, or parent nomination to identify students who may need additional support. Though, these methods are inherently reactive as they track behavior problems after they occur rather than assessing student needs early. Traditionally, students have also been identified for more intensive supports via teacher nomination. However, research suggests that teachers are inaccurate in their nomination procedures and vary in their ability to accurately identify students at risk. Research has indicated that teachers are less likely to refer students with mental health or behavioral concerns versus academic concerns (Walker, Nishioka, Zeller, Severson, & Feil, 2000). Further, teachers are more likely to identify students with externalizing concerns rather than internalizing symptoms (Lane & Menzies, 2005; Richardson, Tolson, Huang, & Lee, 2009; Soles, Bloom, Heath, & Karagiannakis, 2008).

Research suggests using universal screening to identify students at risk for emotional and behavioral issues. Several universal screening tools have demonstrated promising technical adequacy. The Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2015), Behavior Intervention Monitoring Assessment System (BIMAS-2; Meier et al., 2008) and

the Student Risk Screening Scale (SRSS; Drummond, 1994; Lane et al., 2007) are some of the standard options for schools to identify risk for psychopathology in students. The Social Academic and Emotional Behavioral Risk Screener (SAEBRS; Kilgus & von der Embse, 2014) and the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) use a dual-factor theoretical model by showing not only risk for psychopathology but also student strengths in various areas. While there is research to support each of these screeners, the SAEBRS is the only universal screener to date that has met psychometric criteria set forth by the Technical Review Committee of the National Center for Intensive Intervention. Currently, there are no systematic screening procedures incorporated within the Pyramid Model. Students are typically referred by their teacher for intensified intervention. Teacher referral processes often include much less data, if any, that a universal screening procedure would produce, which could guide initial tier 2 intervention and instruction. As such, tier 2 assessment methods are critically important to guide student service teams in selection and implementation of tier 2 interventions.

Tier 2 Assessment. Consistent with intervention research, the research on Tier 2 assessment methods have lagged significantly behind Tier 1 universal screening and Tier 3 individualized assessment practices. As described above, if a student is identified as needing additional supports through universal screening, teacher nomination, or other forms of identification, the student is typically assigned to a one-size-fits-all intervention. The purpose of a Tier 2 intervention is to be efficient, general, and easily integrated into the school's infrastructure to allow continuous availability to students (Hawken et al., 2009). This model of Tier 2 intervention implementation has generally lacked a specific connection to Tier 2 assessment practices. This is unfortunate given research suggests the need for differentiation of Tier 2 supports depending on student needs.

Tier 3 Assessment. Similar to Tier 3 intervention practices, Tier 3 assessment is highly individualized and often resource-intensive. The most common type of assessment on this level is a comprehensive evaluation to determine special education eligibility. Depending on the type of concern, this usually consists of cognitive and/or achievement testing, behavioral or adaptive rating scales completed by the parent, teacher and/or student, developmental or social histories, review of records and previous interventions, and direct observations. For students with behavior concerns, an FBA may also be conducted as a part of this evaluation. Outside of the formal evaluation process, a school may conduct an FBA to help inform a behavioral intervention plan. In preschools, FBA's are typically reserved for severe cases and conducted only at the Tier 3 level due to the time and resource-intensive nature of the data collection. As with FBA's in other settings, an FBA in schools also involves direct and indirect data collection methods. Direct observations usually require highly trained personnel (e.g., school psychologist, behavior analyst) to conduct a systematic direct observation (SDO) or other formalized assessment method. Indirect interviews and rating scales must also be conducted or provided and scored by trained personnel. Following the collection of these data, it is compiled into a comprehensive report. This report leads to a behavior intervention plan (BIP) or behavior support plan, which is informed by the determined function of the student's behavior.

Use of Individualized Assessment at Tier 2

The primary concern of interest in this study is the lack of Tier 2 assessment research and practice guidance, particularly in early childhood settings. Despite the logic and efficiency of using a one-size-fits-all Tier 2 intervention approach for students with behavioral problems, research has suggested Tier 2 interventions are more effective when function is considered (Hawken et al., 2011; Lane, Capizzi, Fisher, & Ennis, 2012; March & Horner, 2002). Currently, within early childhood settings and elementary settings, function is typically not considered until

Tier 3. To tailor Tier 2 intervention procedures to a student's function, a form of assessment must be utilized to inform this adaptation. While this would require somewhat more resources at the Tier 2 level, this type of practice is supported under efforts towards prevention as catering Tier 2 intervention may prevent students from needing more intensive Tier 3 assessments and services. Indeed, more substantial effects have been found with function-based adaptations of these interventions when FBA findings are used (Campbell & Anderson, 2008; Kilgus et al., 2017; Klingbeil, Dart, & Schramm, 2018).

While some researchers have called for the use of FBA's at the Tier 2 level in light of this research (Reinke et al., 2013), others have raised concerns about feasibility (Cheney, Flower, & Templeton, 2008). Indeed, the resources and time necessary to complete an FBA that includes multiple types of data collection and specialized personnel are not feasible for the relatively large number of students who typically require Tier 2 supports. In light of this, some have suggested using a single tool or measure may be appropriate given the lower stakes nature of these decisions (Dunlap et al., 2018; McIntosh et al., 2008). Yet the question remains of which tool would yield the most reliable data while maintaining feasibility for use on a larger portion of students.

Characteristically, direct measures (e.g., systematic direct observations) yield the most reliable data. SDO's involve a third-party observer conducting highly structured observations. This results in highly objective and low inference data but requires an individual who has been trained in such observational techniques and coordination with teachers and staff to conduct multiple observations across settings. As such, this method may not be feasible to implement on a Tier 2 scale, and therefore, some have proposed indirect methods for Tier 2 usage rather than direct observations for feasibility purposes. Indirect functional assessment methods include rating scales (e.g., Functional Analysis Screening Tool; Iwata & DeLeon, 1995) or functional

interviews (e.g., Functional Assessment Checklist for Teachers and Staff; March et al., 2000), which raises the concern of reliability and the inherent subjectivity that accompanies indirect methods through a parent or teacher's perspective. Further, some studies have revealed inconsistent performance of such rating scales. In particular, rater bias is of concern in scales and interviews that ask raters to describe and assess behavior up to weeks or even months in the past (Christ, Riley-Tillman, & Chafouleas, 2009). For example, Iwata, DeLeon & Roscoe (2013) found the correct function was identified using the Functional Analysis Screening Tool only 63.8% of the time across 69 cases.

Direct Behavior Rating

Recently, researchers have begun investigating the extent of appropriate usages of direct behavior rating (DBR) methodology. DBR is an assessment method that involves instructional staff (i.e., a teacher) observing a student and estimating the rate of disruptive behavior. DBR is a synthesis between the feasibility of rating scales and the reliability of direct observations given the low latency between observed behavior and rating (Chafouleas, 2011). DBR methods have demonstrated accuracy and validity compared to true scores, typically measured through a Systematic Direct Observation (SDO; Kilgus, Riley-Tillman, Stichter, Schoemann, & Bellesheim, 2016; Miller, Chafouleas, Riley-Tillman, & Fabiano, 2014). One study found DBR methods to be highly correlated (within 1-2 points) of an SDO (Riley-Tillman, Chafouleas, Sassu, Chanese, & Glazer, 2008) and has also shown stability in score estimates over time (Kilgus et al., 2016).

However, past research has primarily focused on rating the degree of presence of disruptive behavior (e.g., (Kilgus et al., 2016; Riley-Tillman et al., 2008). Some researchers have started evaluating the utility of direct behavior rating methodology in assessing not only the occurrence of disruptive behavior but the consequences of that behavior as well (e.g., Kilgus et

al., 2017). Conditional probability estimates are often used to determine probable functions of behavior by determining how often a specific consequence follows disruptive behavior (Martens, DiGennaro, Reed, Szczech, & Rosenthal, 2008). As such, recent investigations into the utility of DBR to determine the function of behavior have been based on similar conditional probability estimates. Specifically, raters are asked to rate the percentage of time that a student receives a consequence (e.g., adult attention) out of the total time of disruptive behavior.

A study conducted by Kilgus et al. (2017) trained undergraduate students to accurately rate student behavior from video clips using a new tool called the *Intervention Selection Profile-Function* (ISP-Function). In this study, undergraduate students were able to assess the disruptive student's behavior and consequences within 0 to 2 points of the correct score, determined by an SDO (Kilgus et al., 2017). However, a similar study conducted by South (2017) used a small sample of paraprofessionals and found they were able to accurately rate the amount of disruptive behavior and the conditional probability of adult attention only, not other consequences. This may be the result of floor effects, as few students received any consequences other than adult attention. Further, some research has suggested that DBR data may be more reliable when collected by a classroom teacher rather than a paraprofessional (Johnson et al., 2016). These preliminary findings in the use of DBR to evaluate the potential function of behavior are promising but inconsistent and warrant further investigation.

This study intends to evaluate a new form of FBA developed with DBR methodology, used in a previous study conducted by Kilgus et al. (2017), the *ISP-Function*. The ISP-Function is a one-page DBR form intended to be used by teachers to assess the function of behavior. The tool is completed in two steps. The first step is to estimate the percentage of time across the whole interval that the student engaged in disruptive behavior. The unipolar graphic rating scale is 0 to 10, representing 0% to 100% of the time. The second step is to rate each of the four

consequences of (a) adult attention, (b) peer attention, (c) escape or avoidance, and (d) access to tangibles or activities. These ratings are also rated on a unipolar graphic rating scale from a scale of 0-10. This tool is meant to enhance preventative services by allowing for individualization of less intensive intervention to prevent the need for more intensive Tier 3 interventions. Promising results have been found in beginning work with this tool in elementary settings. One correlational study evaluated the temporal reliability, validity, and accuracy of this measure. Kilgus et al. (2019) found support for strong temporal reliability, validity, and accuracy compared to systematic direct observation. However, these results varied somewhat across ISP targets. Together, results indicate initial support for the usage of the tool to assess the amount of disruptive behavior and degree to which the behavior is met with attention from others. However, this study's somewhat inconsistent results implicate the need for future research to evaluate the type and level of training necessary for teachers to use the ISP-Function accurately. Further, this tool has not yet been evaluated for use in preschool settings.

Advancing Behavioral Assessment Practices in Early Childhood

There has been a growing focus on behavior problems in toddlers and preschoolers as the field has been turning towards early intervention. Common behaviors seen in early childhood populations include noncompliance, tantrums associated with emotion dysregulation, aggression, property destruction, and self-injury (Wakschlag et al., 2007). While not all children who display these behaviors require professional attention, between 8-17% of preschool children demonstrate clinical severity of disruptive behavior (Briggs-Gowan, Carter, Skuban, & Horwitz, 2001; Egger & Angold, 2006; Furniss, Beyer, & Guggenmos, 2006; Lavigne, LeBailly, Hopkins, Gouze, & Binns, 2009).

Some research has indicated that children from lower income families are at increased risk of behavior problems (Walker, Nishioka, Zeller, Severson, & Feil, 2000). It is estimated that

this difference may be due to their higher risk of stress factors, limited access to quality childcare and education, and the effect of stress on positive parenting practices (Qi & Kaiser, 2003; Linver, Brooks-Gunn, & Kohen, 2002). To close the achievement gap between low-income students and their wealthier peers, educators and federal and state leaders have begun to focus more on early learning, indicators of school readiness and the importance of early intervention (Blair & Raver, 2015; Blair, Fox, & Lentini, 2010; Reardon, 2013). Specifically, kindergarten and school readiness have become a national priority (Bowman, Donovan, & Burns, 2000).

Many tools have been developed to screen and assess deficits in early childhood. Table 1 is a review of the most commonly used tools in early childhood settings. While many have been designed specifically for the early childhood population, some of these tools lack adequate investigation of the tool in children under the age of 5 (e.g., Strengths and Difficulties Questionnaire, Pediatric Symptom Checklist). Most tools are designed to be completed by parents of the child and are most often used in pediatric or primary care settings. Of the reviewed tools, only the Sutter-Eyberg Student Behavior Inventory-Revised (Eyberg & Pincus, 1999), The Preschool and Kindergarten Behavior Scales-Second Edition (PKBS-2; Merrell, 2003), Strengths and Difficulties Questionnaire (Goodman, 1997), Devereux Early Childhood Assessment (LeBuffe & Naglieri, 1999), Early Childhood Behavior Screen (ECBS; McCarney, 1994) and the Early Childhood Inventory-4 (ECI-4; Gadow & Sprafkin, 1997) are developed for use in the schools. The purpose of these tools is to detect the presence of developmental delays, behavioral difficulties, and social-emotional deficits at a clinical level but do little to inform intervention or the next step in practice. A notable exception is the Early Childhood Behavior Screen, which does offer a series of pre-developed goals, objectives, and interventions through their Early Childhood Behavior Intervention Manual. Though, these are individualized to the problem (e.g., reinforce student for working independently to promote independent working) rather than

personalizing strategies to the student's function (e.g., ensure peer attention for working alone through a group contingency reinforcement system).

Alternatively, most functional assessments (e.g., Functional Assessment Checklist for Teachers and Staff) have not been examined in early childhood settings. Rather, some programs such as the Pyramid Model (Fox, Dunlap, Hemmeter, Joseph, & Strain, 2003) and Prevent-Teach-Reinforce for Young Children (Dunlap, Wilson, Strain, & Lee, 2013) utilize practitioner-designed direct behavior ratings to assess and progress monitor specific behaviors and their frequency, intensity, or duration. These tools are highly feasible as they are catered to the student's displayed behavior and goals of the intervention team. One study that evaluated the usage of DBR in preschools to assess social behavior (i.e., working to resolve conflict and interacting cooperatively) found that ratings varied considerably between the behavior that was being measured as well as the rater (Chafouleas, Christ, Riley-Tillman, Briesch, & Chanese, 2007). These results indicate the need for additional investigation in the use of DBR in preschool populations and suggest rater training may be necessary for reliability.

Indeed, there are currently no validated brief functional assessment tools at the preschool level. Instead, to determine function, a full functional behavior assessment is typically conducted. Due to the time and resource needs of a complete FBA, this is generally reserved for students with severe mental health needs at the Tier 3 level. Consequently, students at the Tier 2 level, do not receive any individualization as the standard procedure for determining function is too resource-intensive. This is a critical gap in the literature as research has shown that individualizing intervention to a student's specific function is more effective than a one-size-fits-all approach (Gage et al., 2012; Walker, Chung, & Bonnet, 2018). The ISP-Function allows teachers to briefly assess and determine a probable function at the Tier 2 level using minimal resources. This information can be used by student intervention teams to tailor basic Tier 2

intervention (e.g., Check In, Check Out) to the student's function to enhance effectiveness and ultimately prevent students from reaching a Tier 3 level of need.

Table 1: Common Early Childhood Assessment Tools

Measure	Ages (in years)	Rater	Target	Function
Ages and Stages Questionnaire (ASQ) - 3rd Ed.	0-5	P	D	No
Behavioral Assessment of Baby's Emotional and Social Style (BABES)	0-3	P	B, SE	No
Brief Infant Toddler Social-emotional Assessment [BITSEA] (2005)	1-3	P	SE	No
Child Behavior Checklist (CBCL)	1.5-5	P	B	No
Children's Behavior Questionnaire	3-7	P	B	No
Devereaux Early Childhood Assessment	2-5	P, T	B	No
Early Childhood Behavior Screen	3-5	T	B	No
Early Childhood Inventory-4 (ECI4)	3 to 5	P, T	D	No
Eyberg Child Behavior Inventory (ECBI) and the SutterEyberg Student Behavior Inventory-Revised (SESBIR)	2-16	P, T	B	No
Parents' Evaluations of Developmental Status (PEDS)	0-8	P	D	No
Pediatric Symptom Checklist (PSC)	4-16	P	SE	No
Pervasive Developmental Disorders Screening Test-II (PDDST-II), Stage 1-Primary Care Screener	1-4	P	SE	No
Preschool and Kindergarten Behavior Scales – 2nd Edition (PBKS-2)	3-6	P, T	B	No
Strengths and Difficulties Questionnaire (SDQ)	3-16	P, T	B	No
Temperament and Atypical Behavior Scale (TABS screener)	1-5	P	D	No

P=Parent/Caregiver; T=Teacher; B= Behavior; D= Developmental; SE= Social-emotional

Teacher Training in Behavior Assessment Methods

Few rating scales or assessment methods in schools utilize rater training due to limited resources. However, rater training has been shown to improve the psychometric adequacy and the data-based decisions made based on these assessment tools (Chaflouleas, 2011). The issue of rater training is particularly relevant for the ISP-Function in early childhood education. This population is acutely important given the variety in training backgrounds of early childhood educators. The type of early childhood education center (e.g., preschool, Headstart) dictates the credentials necessary to teach in early childhood ranging from a high school diploma to a

bachelor's degree in early childhood education. This wide range of training backgrounds makes a unique landscape for examining the teacher's role in conducting direct behavior ratings. One of the primary strengths of this tool is the utilization of the teacher in collecting DBR data instead of using resources necessary in many other tools. However, to gather quality data through teacher observation, the adequacy of this tool relies on teachers' ability to collect accurate data. A review of the issues and research around DBR by Chafouleas (2011) outlined the recommendations based on these inconsistent findings. The review poses that due to individual differences, some level of training should be provided to improve outcomes, accuracy, and reliability of ratings. The recommendations further state that the trainings do not need to be high in intensity and that (1) clear definitions and modeling of rating, (2) practice and feedback, and (3) behavioral examples across the scale range are likely to be most effective components of rater training. These recommendations are consistent with best practices of professional development and teacher training the suggests utilizing a combination of theory and providing opportunities for practice (Barnes & Harlacher, 2008). Joyce & Showers' (2002) seminal research on staff development also outlined observation, practice, and feedback as the critical three components of professional development that promotes teacher change.

Several studies have evaluated the utility and effectiveness of trainings for teacher collected DBR methods in the past. A study from Chafouleas et al. (2012) assessed differences between a standard training protocol with 3-6 practice ratings and feedback, a protocol that included frame of reference training or a comprehensive protocol that included frame of reference training and rater error training. The results from this study suggested that the full training package did improve the accuracy of rating academic engagement and disruption. However, this study also found that a comprehensive package may not be significantly more beneficial than a standard training package. A study conducted by LeBel and colleagues (2010)

compared an indirect (i.e. recorded) training protocol and a direct training protocol with performance feedback to no training. The results suggested that neither the indirect or direct teacher training improved the accuracy of teacher ratings. These inconsistent findings of the effectiveness of rater training in DBR indicate at least some level of inter-individual differences in levels of necessary training (Chafouleas 2011).

The novelty of the ISP-Function from other DBR rating methods is the addition of the consequence items that prompt teachers to observe and record the consequences of behavior and the actual behavior to determine a potential function of the behavior. This necessitates the question of how much training is necessary for a teacher to conduct functional-based assessments. One study that looked at early childhood educators' base knowledge of FBA's and behavioral principles suggested preschool teachers and early interventionists alike may benefit from additional training of functional behavior assessments (Martin, 2016). For effective implementation of this tool, it is critical that teachers understand basic principles of behavior and consequences, as this is what they will be looking for during the observations. Some research has evaluated teachers' knowledge of functional behavior assessment processes and supports in students. Hesney (2011) evaluated 108 general education and special education teachers' knowledge of FBA's. The sample included teachers from preschool-high school level. There was no significant difference between general education and special education teachers in their knowledge of FBA's. The average score on the knowledge assessment was 7.37 out of 12, indicating more training around FBA's for teachers may be necessary. Further, Kircher (2009) assessed 87 teachers' knowledge of the FBA process and found that the majority of participants did not have a proficient understanding of the process. Out of the eight questions used to assess knowledge of FBA's, none of the participants accurately answered all questions, and ten teachers got every question incorrect. Both of these studies used a limited number of preschool teachers,

so to build on this research, Martin (2016) studied the perceptions, self-efficacy, and knowledge of FBA's in preschool teachers. Martin (2016) found that both preschool teachers and early childhood interventionists lacked critical knowledge regarding functional behavioral assessments. This research suggests the need for increased training in the FBA process, particularly in the preschool teacher population (Martin, 2016). Teacher training in this project refers to any instruction or support provided to teachers. Several types of teacher training were evaluated that include different elements of instruction, including basic exposure or practice, and performance feedback.

As aforementioned, the ISP-Function has yet to be examined in a preschool population. This tool was initially validated in a study of 213 undergraduate student raters (Kilgus, Kazmerski, Taylor, & von der Embse, 2017). A follow-up study published in 2019 by Kilgus et al. evaluated the reliability, validity, and accuracy of scores from the ISP-Function. This study included 34 student-teacher dyads and further supported the use of this tool. It found teachers were generally able to accurately rate students with disruptive behavior within 0-2 points of the true score (Kilgus et al., 2019). More detailed findings of these studies can be found in the Measures section below. In both of these studies, the research team used a comprehensive training package that included basics of behavioral principles and performance feedback. Given that this training package resulted in strong reliability and accuracy of ratings in both undergraduate student raters (Kilgus, Kazmerski, Taylor, & von der Embse, 2017) and in elementary school teachers (Kilgus et al., 2019), this study modeled the components after this comprehensive package.

Summary

Teachers in early childhood settings continue to struggle with addressing behavioral concerns (Kaufmann & Wischmann, 1999). Despite the strong foundations created by programs

such as the Pyramid Model and PTR-YC for Tier 1 and Tier 3 practices, more work is necessary to evaluate intervention and assessment research for students with less severe behavior problems served at the Tier 2 level. Functional-based interventions have shown to be more effective than interventions that are not informed by function. Yet, functional assessments are often conducted only for a small percentage of students with the most severe behaviors given the time and resource-intensiveness typically associated with FBAs. Using a single functional assessment tool rather than a comprehensive assessment and observation process may be a feasible way to address the function of a student's behavior prior to needing Tier 3 services. Direct Behavior Ratings offer a feasible, low latency, and high objectivity way for teachers to rate student behavior. However, these practices in early childhood settings are particularly challenging, given the ranges of teacher experience, qualification, and training in behavioral principles. Therefore, the question of the degree of teacher training necessary to implore the use of this assessment strategy is critical. The following chapter describes the methods used to evaluate the necessary levels of early childhood teacher training in using the direct behavior rating tool, the ISP-Function.

Chapter III: Method

This chapter describes the study design, participants, and procedures for the current study. This study was conducted using a concurrent multiple baseline single-case design (Kazdin, 1982). A multiple baseline design was selected due to its systematic ability to detect changes in the dependent variable as a direct result of the independent variable with staggered treatment start points. Further, this design was selected over other single-case methodologies (e.g., ABAB) due to the impossibility of withdrawing the treatment (i.e., teacher training) given the treatment is an acquisition of knowledge. Using a concurrent multiple baseline single-case design, this study evaluated changes in the accuracy of teacher-collected ISP-Function data as compared to a systematic direct observation.

This study used a pre-selected number of baseline and intervention phase data points for each baseline condition as opposed to a response-guided method involving observing the data and implementing the intervention phase once data points are stable (Kratochwill & Levin. 2010). This decision was made based on the feasibility of communicating time commitments to participants and preventing attrition through prolonged data collection.

Research Questions

To evaluate the utility of the ISP-Function by preschool classroom teachers, this study firsts seeks to answer three critical, interrelated research questions:

1. To what extent does teacher training improve the accuracy of ISP-Function data collected?
2. Which training component is the most effective method of training on the accuracy of ISP-Function data collected?

3. Is there a difference between the order in which the training components are implemented on the accuracy of ISP-Function data collected?

The researcher hypothesized that teacher training is necessary to reach higher levels of accuracy in ISP-Function scores. Further, the researcher hypothesized that performance feedback would be the most effective training component alone but that the order of Behavior Basics then Performance Feedback would yield the highest accuracy ratings.

Participants and Setting

Standards set forth by What Works Clearinghouse state that in order to demonstrate an effect of the intervention, at least three demonstrations of the effect must be established. In terms of a multiple baseline study, this is represented as three participants showing an effect, or response to the intervention, in this case, teacher training. More demonstrations of effect increase the confidence in experimental control (Kratochwill & Levin, 2010). This design used three baseline conditions wherein participants in each baseline condition received their first training after the fifth, seventh, and ninth baseline conditions, respectively. This is consistent with multiple baseline methodology of staggering initiation of treatment conditions to reduce the threat to the internal validity of history (e.g., an outside factor that may influence the dependent variable). Additionally, because there are two treatment conditions (Behavior Basics and Performance Feedback), trainings were counterbalanced among participants. To achieve this, two participants were randomly assigned to each of the three baseline conditions. One participant received Behavior Basics training first, and the other participant received the Performance Feedback training first to eliminate ordering and practice effects. These procedures are described in greater detail below. As a result, this study recruited six participants, two for each of the three baseline conditions, in order to counterbalance treatment conditions (see Table 2). This sample

size also allowed for potential attrition of participants while still protecting the ability to potentially demonstrate three intervention effects as required by WWC Standards.

This study took place in six different preschool classrooms. Participants included six preschool teachers and one student with disruptive behavior from each classroom. Teacher inclusion criteria included that the teacher is (1) the primary or co-teacher in a 4-5-year-old class, and (2) expresses interest in participating in the study. Teachers were not excluded based on education or years of experience. This was to replicate the actual sample of early childhood educators who have vastly varying training backgrounds. The inclusion of all individuals who serve the role of the lead teacher will improve the generalizability of these results compared to selecting only those, for example, with bachelor's degrees or higher. Therefore, the only inclusion criteria for teachers was that they are a lead teacher in a preschool (4-5-year-old) classroom at a participating school.

Student inclusion criteria included (1) displays observable disruptive behavior that (2) occurs at least five times per day. The research team used a modified version of the Functional Assessment Checklist for Teachers and Staff (FACTS) to establish these inclusion criteria. A more detailed description of the screening procedure is described in the recruitment section below. The behavior's frequency was to ensure that the disruptive behavior is not a high intensity, low frequency (e.g., tantrum once per week) behavior that would make it difficult to observe on a regular basis. Each student acquired parental consent to be observed by researchers. No participants were excluded based on gender, race, ethnicity, or other factors not addressed in the inclusion criteria.

This research was funded by the National Association of School Psychologists Graduate Student Research Grant (GSRG). All funds received through the GSRG went towards participant compensation and costs incurred for materials related to the project. Each teacher participant

received a $100 gift card incentive for his or her participation. Students were not aware of their involvement (see consent procedures in the Recruitment section) and therefore did not receive any direct compensation. However, the data collected on each student was compiled into an individualized behavior intervention plan report. This report may help to inform practices used to support the student's disruptive behavior. Given that the student and family were not asked to do anything above and beyond what is required of their typical school day, the minimal risk nature of the study, and the potential benefit of the behavior plans yielded from the data collected in the classroom, this research does not violate any ethical standards and does not pose any risk beyond what is present in everyday schooling.

Table 2: Baseline Conditions

Condition	Baseline Points	Intervention Phase 1	Intervention Phase 2
Condition 1	5	TP1: Behavior Basics	TP2: Performance Feedback
Condition 2	5	TP2: Performance Feedback	TP1: Behavior Basics
Condition 3	7	TP1: Behavior Basics	TP2: Performance Feedback
Condition 4	7	TP2: Performance Feedback	TP1: Behavior Basics
Condition 5	9	TP1: Behavior Basics	TP2: Performance Feedback
Condition 6	9	TP2: Performance Feedback	TP1: Behavior Basics

Measures

Intervention Selection Profile-Function (ISP-Function). As described in the previous chapter, the ISP-Function is founded upon direct behavior rating methodology. The ISP-Function measures both the extent to which a student engages in an operationally-defined problem behavior during a pre-specified period (e.g., math instruction, 10:30-11:15) and the frequency with which the behavior is met with four consequences: Adult Attention, Peer Attention, Escape/Avoidance, Access to Tangibles or Activities (See appendix A). This tool was initially validated in a study with 213 undergraduate students to determine if it could be used to collect functional assessment data (Kilgus, Kazmerski, Taylor, & von der Embse, 2017). Participants in

this study were able to generate ratings of a video clip within 0-2 points of the true score, as determined by SDO. A follow-up study published in 2019 by Kilgus et al. evaluated the reliability, validity, and accuracy of scores from the ISP-Function. This study included 34 student-teacher dyads and compared ISP-Function teacher ratings to research assistants' systematic direct observation ratings (SDO). Analyses suggested adequate temporal reliability for assessing adult and peer attention (≥.70) with three observations. Alternatively, when assessing disruptive behavior, escape/avoidance, and access to items or activities, 8-18 observations would be required to reach adequate reliability (Kilgus et al., 2019). Similarly, analyses suggested accurate mean ratings with statistically significant correlations for adult attention (r = .65), peer attention (r = .55), as well as the disruptive behavior (r = >.62). However, correlations were lower for access to items and tangibles/activities (r = -.15) and escape/avoidance (r = .22). Finally, while Kappa values suggested that agreement between ISP-Function and SDO scores were poor to fair in accuracy compared to SDO, difference scores indicated that teachers were able to generate ratings within 0-2 points of true scores while engaged in instructional actives. (Kilgus et al., 2019).

Systematic Direct Observation. The criterion measure used in the observations was a systematic direct observation (SDO) tool. SDO is often considered the gold standard of behavior assessment and one of the most common tools used within the functional behavior assessment process (Lloyd, Weaver, & Staubitz, 2017). Further, it is often used within the literature and as a robust psychometric foundation (Chafouleas, Kilgus, Riley-Tillman, Jaffery, & Harrison, 2012). It is also commonly used as a part of comprehensive functional behavior assessments (FBA) in schools. Observations using the SDO form utilize behavior-consequence (BC) partial interval recording. Accordingly, an interval was marked if disruptive behavior occurred within the interval. Two separate scores were calculated for each of the five targets. First, the total

frequency across the entire observation period was calculated for each target. Following, the percentage of intervals in which each target occurred was recorded. Each consequence corresponds to item 2-5 on the ISP-Function (see appendix B).

Implementation Fidelity Checklist. A fidelity checklist was used during each training to ensure each component of the training is completed. Immediately following each training, the primary researcher completed the checklist. The checklist was developed by the research team and included all critical elements of each training. The fidelity checklist corresponded directly to the training protocol sections. A copy of each fidelity checklist can be found in Appendix D.

Procedures

Research Design. In order to answer the research questions, the researcher used a multiple baseline single-case design. By using a concurrent multiple baseline design, the researcher was able to evaluate whether changes in ISP-Function accuracy scores are related to the implementation of various teacher training components. The researcher concurrently evaluated all six participants using six baseline conditions (see Table 2). The procedures for recruitment, data collection, and analysis are described in greater detail below. The independent variable in this study is the training protocol. Below each training protocol is described in greater detail. This study also has three dependent variables: Overall Agreement (OA), Disruptive Behavior Agreement (BA), and Consequence Agreement (CA). Agreement is calculated by calculating a difference score between the Systematic Direct Observation that is completed by the researcher, and the ISP-Fx, which is completed by the teacher. Disruptive Behavior Agreement (BA) is the absolute value of the difference score on the first item of the measure, which asks the teacher to estimate the percentage of problem behavior across the interval. The teacher rates this item on a scale from 0-10 and therefore the range for BA is 0-10. Consequence Agreement (CA) is the combined absolute value of the difference between raters on all four

consequence items. Each item is rated on a scale from 0-10, which results in a combined range of 0-40 for the variable. Overall Agreement (OA) is the combined absolute value of the difference score for all five items on the ISP-Fx measure. The combined range of scores for this variable is on a scale of 0-50. All graphs depict the full range of the possible difference scores. On each graph, the range on the y-axis is equal to the total possible range of disagreement for the variable.

Recruitment. The researcher recruited local preschools in the Tampa Bay Area. Once the school administration agreed to participate in the study, the principal investigator recruited individual teachers and discussed potential students that may fit the student inclusion criteria. Given the importance of observing disruptive behavior during the interval and to avoid the occurrence of invalid data, the research team implemented a screening procedure to maximize the possibility of observing disruptive behavior. A modified version of the Functional Assessment Checklist for Teachers and Staff (FACTS; see appendix C) was utilized to determine appropriate student participants with the teacher. To pass the screening, the teacher had to indicate that the problem behavior occurred every day, at least five or more times per day. Further, within the identifying routines and likelihood of problem behavior, at least one activity during the day had to be rated a 5 or 6, meaning it has a high probability of occurring during that time. This was intended to ensure the behavior was predictable enough to schedule an observation during a high likelihood time and avoid situations in which disruptive behavior was episodic or unpredictable. Of the fourteen students screened, eight students passed the screening. Two of those did not receive parental consent, and the remaining six were the study participants.

The target behavior of all of the students who participated was generally defined as class disruption and included behaviors such as calling out or being out of area. It also included more severe behaviors such as screaming, aggression, throwing self to the floor, rolling around, and

throwing items. Most of the observed behavior mild in intensity (e.g. calling out, out of area). Teachers overwhelmingly rated morning circle time to be the most problematic time of the day and the most likely time for behavior to occur. As such, most observations occurred during morning circle time (approximately 9:00 AM) since observations were scheduled during the most problematic time of the day, according to ratings.

If an interested teacher had a student that meets inclusion criteria, parental consent was sent home. Once parental consent was obtained, the teacher met with the researcher to discuss and sign the informed consent to participate in the study. After both teacher and parental consent were received, the teacher-student dyad was enrolled in the study. If parental consent was not returned or was denied, the student was not eligible for participation. The teacher could then select another student to be screened, or the teacher was removed from the study sample. Only two parents did not consent to participate in the study, and those teachers did not participate in the study.

Once enrolled, the researcher met briefly with the teacher to discuss the procedures and logistics of the study and the target student. A time was set (e.g., before school) to conduct each of the trainings. The training dates occurred following the pre-selected number of baseline data points collected. The FACTS completed during the screening process served as a guiding tool to determine when the observations would take place. In all instances, observations were scheduled during a time of day when the behavior was rated a 5-6 on the FACTS (very likely to occur). This was to maximize the probability of observing the disruptive behavior. All efforts were made to have the observations occur at approximately the same time for each observation period.

To decrease the threat of history to internal validity, randomization is often used (Kratochwill et al., 2010). In this study, once a teacher-student dyad was enrolled, they were randomly assigned to a baseline condition using an online randomization tool (see Table 2 for all

conditions). The baseline conditions were distinguished by the number of baseline points collected for the participant and the counterbalancing of training protocols. The counterbalancing of training protocols was to protect against order effects and to determine the specific components of the training that are most effective and necessary to achieve accuracy. There were two participants in each core baseline condition (as illustrated in Table 2). Two participants received the training after five baseline observations, two received the training after seven baseline observations, and two received the trainings after nine baseline observations. Of the pairs in the same baseline condition, one received the behavior basics protocol first, and the other received the performance feedback protocol first. The various phases, including baseline and descriptions of the training, are described below.

Baseline Condition. The baseline phase consisted of basic exposure to the measure. The researcher briefly explained the measure and provided few details. This introduction to the tool took no longer than 3-5 minutes. The researcher briefly explained the two steps of the measure. The first step is to rate the percentage of disruptive behavior that occurred during the observation on a scale of 0-10. The second is the rate each of the four consequences based on how often the disruptive behavior was followed by each on a scale of 0-10. This brevity was to represent current practices in schools, which commonly consists of little to no training of new measures used. Following basic exposure, the predetermined number of observations took place based on the participant's randomly assigned baseline condition. The observation procedures are described below.

Training Protocols. As aforementioned, two training protocols were counterbalanced. The first was referred to as 'Behavior Basics,' and the following was referred to as 'Performance Feedback.' Both trainings were conducted on an individual basis with just the teacher and the researcher at the teacher’s school. The researcher used a computer to display a PowerPoint

presentation that went through the training. The trainings occurred at a time convenient for the teacher either before school, after school, or during a teacher break during the day (e.g., naptime). The trainings were developed to be easily understood by teachers who may have little to no training in behaviorism or behavioral principles. The following are detailed descriptions of the two trainings.

The Behavior Basics protocol introduces the teacher to basic behaviorism principles, including an introduction to the effects of antecedent conditions or events and consequences on the continuation of problem behavior. The definition and concept of 'function' are described as a consequence that maintains a behavior. Following this, the teacher is introduced to four of the most common behavior functions in schools, which are the functions on the ISP-Function (i.e., Adult Attention, Peer Attention, Escape or Avoidance, and Access to Tangibles or Activities). It is explained that consequences do not need to be intentional or positive to serve as a function. For example, reprimanding a child still serves as adult attention. Further, the child does not need to be aware of the function of their behavior. Age-appropriate examples are given throughout to illustrate these concepts. The training also describes the importance of function on intervention planning and treatment of behavior problems in the classroom. The teacher is taught that for effective behavior change to occur, the reinforcing consequence must be removed following undesirable behavior and must be offered for positive behavior. Teachers were encouraged to ask questions throughout, and the researcher provided the answers. This is in contrast to the basic exposure procedure where few questions were answered to simulate current practice in schools. This training typically took a total of 10-15 minutes to complete.

The performance feedback protocol included opportunities for teachers to receive performance feedback. This is a method commonly used in training that allows participants to practice the skill (i.e., using the tool) and receive feedback on how well they demonstrated it.

This allowed an opportunity for participants to practice and adjust incorrect behavior for future demonstrations of the skill. In this training protocol, teachers viewed two pre-recorded videos of disruptive behavior. The videos have graduate students simulating students in a classroom. Teachers were instructed to watch one particular student displaying disruptive behavior. After watching the approximately three-minute-long video, teachers were asked to rate the student's behavior and the consequences that followed using the ISP-Function. The teacher followed the two steps of the measure: (1) rate the percentage of time the student was disruptive on a scale of 1-10 and (2) rate the percentage of disruptive behaviors that were met with each of the four consequences (adult attention, peer attention, escape, access to tangibles or activities). After rating the student's behavior, the trainer reviewed the correct answers with the participant and answered any questions. The correct scores were predetermined using systematic direct observation. After completing the performance feedback practices for both videos, the training concluded. Again, teachers were encouraged to ask questions throughout the training. This training typically took between 10-15 minutes to complete. To ensure the teacher training's implementation integrity, each training protocol included an integrity checklist to ensure that the researcher included each component of the training (see appendix D).

Observations. Following the baseline condition, the researcher conducted between five and nine observations (based on the participant's randomly assigned baseline condition). After each training, the researcher conducted five observations in the classroom. The researcher gave the teacher the ISP-Function form and signaled the teacher when the 10-minute observation period began. The teacher continued with typical instructional activities but maintained some extra attention to the target student. The researcher sat in the classroom and completed a systematic direct observation (SDO) for 10 minutes on the target student. Following the observation, the teacher filled out the ISP-Function based on the behavior observed in the

observation period. This typically took no more than 30 seconds. The SDO form and the ISP-Function yield similar results that could be easily compared for accuracy. Specifically, the SDO can be used to calculate the total percentage of time the student was disruptive and conditional probabilities of the consequences following the behavior.

Using the SDO as the true score, each data point was calculated as an accuracy rating for each of the five targets. The SDO produces the percentage of time the student was disruptive. The SDO also allows the scorer to calculate the conditional probability (number of times the disruptive behavior was met with each consequence divided by the total number of disruptive instances) in the form of a percentage. These percentages were divided by 10 to match the metric used on the ISP-Function. These values were subtracted from the ISP-Function scores, and the absolute value of the difference between the SDO calculation and the ISP-Function scores was documented as the accuracy score. The formulas for X_{DB} (accuracy score for disruptive behavior) and X_{AA} (accuracy score for adult attention) are shown below, where I_{DB} = the number of intervals of disruptive behavior and F_{AA} = frequency of adult attention and F_{DB} = the frequency of the disruptive behavior. All other consequences were calculated using the same formula as used with adult attention with the appropriate scores.

$$X_{DB} = \left| \left[10 \left(\frac{I_{DB}}{30} \right) \right] - ISP_{DB} \right|$$

$$X_{AA} = \left| \left[10 \left(\frac{F_{AA}}{F_{DB}} \right) \right] - ISP_{AA} \right|$$

Missing Data. In the event of absences by either the child or the teacher (or another event which makes data collection impossible), which resulted in missing data, several decision rules were established. For three missing data points, additional days were added to the end of data collection and, therefore, did not influence the total number of data points collected or the

number of data points in either phase. None of the participants experienced excessive absences over three, and so there were no missing data.

Analysis

The preliminary plan for data analysis included three primary steps using five tests. First, the integrity of the data was evaluated through interobserver agreement and implementation integrity using fidelity checklists. Next, the researcher conducted visual analysis consistent with What Works Clearinghouse standards for single-case design studies (Kratochwill et al., 2010). Finally, the researcher conducted both a non-parametric and parametric effect size test using the Tau-*U* statistic (Parker, Vannest, Davis, & Sauber, 2011), a non-overlap index, as well as an effect size through multi-level modeling (Ferron, Bell, Hess, Rendina-Gobioff, & Hibbard, 2009).

Interobserver Agreement. To ensure the reliability of the observational data points, the interobserver agreement was calculated. What Works Clearinghouse standards require interobserver agreement for at least 20% of all observational data points. The research team anticipated barriers in obtaining this level of interobserver agreement, and therefore, the team participated in interobserver calibration before data collection. The primary researcher trained all research assistants who conducted observations and facilitated observation calibration. The data collectors watched videos of students with disruptive behavior and rated them using the Systematic Direct Observation form (see Appendix B). The team watched the videos of classrooms and independently rated a student in the videos. Following the video, the team would review scores and discuss any discrepancies. The team continued until all observers in the team consistently rated students with at least 90% reliability. To further ensure the reliability of the observational data points, the interobserver agreement was calculated for 12% of observations. Agreement (IOA) was calculated directly following the observation by comparing both

observers' observation forms. Agreement (IOA) was calculated directly following the observation. If the observation did not obtain at least 90% agreement, the observation was repeated. This did not occur in any of the IOA observations. The agreement was calculated through a point to point ratio using the following equation, where A is the number of agreements, and D is the number of disagreements:

$$IOA = 100\,[\frac{A}{A+D}]$$

Implementation Integrity. The integrity of the trainings was determined using fidelity checklists. The checklists were a self-assessment by the researcher that conducted the training. The checklist had each component of the training that needed to be completed. A percentage of the components were calculated for each training session using the following equation, where CC = number of completed components and TC = the total number of training components:

$$Integrity = 100(\frac{CC}{TC})$$

Visual Analysis. Visual analysis is traditionally used in single-case design studies to determine overall effects (Barlow et al., 2009; Kazdin, 1982). The researcher followed the guidelines set forth by What Works Clearinghouse (WWC; Kratochwill et al., 2010), which consists of four steps. First, the stability of baseline data points was analyzed. Next, the data within each phase was analyzed to assess predictable patterns within phases. Third, a comparison was made between baseline and intervention phases to evaluate if changes in teacher accuracy in using the ISP-Function are tied to the implementation of the rater training. Finally, information from all participants was integrated to determine overall demonstrations of effect. At least three demonstrations of the intervention effect and no non-effects indicate strong evidence of a causal relationship. Moderate evidence of a causal relationship is indicated by at least three demonstrations of the intervention effect and at least one non-effect. There is no evidence of a

causal relationship if there are fewer than three demonstrations of an intervention effect. Also, the level (e.g., mean), the trend (e.g., slope), variability (e.g., deviation from the trend), the immediacy of effect, overlap, and consistency of data patterns were considered (Kratochwill et al., 2010).

Effect Size. An effect size was calculated using an evaluation of overlapping data points using the Tau-*U* statistic (Parker et al., 2011). This method pairs each baseline observation point with each intervention phase observation to determine the number of positive (P), negative (N), and tied (T) pairs and calculates the percentage of data point pairs that overlap from baseline to treatment phases. In addition to conducting a non-overlap effect size, a parametric effect size was also calculated using multi-level modeling (Ferron et al., 2009).

Ethical Considerations

All procedures were submitted and approved by both the University of South Florida Institutional Review Board (IRB; see appendix E for IRB approval letter) and the participating school's IRB, if applicable. As described in the recruitment process, the teacher's initial interest prompted the researcher to meet with the teacher and select an appropriate student in their classroom. Once a student was selected, a parental consent form was sent home (see appendix F). Parental consent must have been received before any research procedures begin. Assent was not collected due to (1) the young age of the child and (2) potential observer effects if the child was aware that the teacher and researcher were observing his or her behavior. This is a typical procedure for studies that use observational data. Following the receipt of parental consent, teachers met with the researcher to review and sign an informed consent form (see appendix G). It was emphasized that participation is purely voluntary, and data collected from the study was not utilized for any performance reviews or influence their employment in any way. Further, the

parental consent form emphasized the voluntary nature and participation would not affect the services the student receives in school or academic standing of the student.

All identifying data collected was kept confidential. All data were de-identified using assigned research participant numbers. The researcher and approved research personnel were the only individuals with access to the data set prior to de-identification. All data and identifying information were stored in a password-protected file and will be kept for five years as dictated by USF IRB protocol. Finally, in single-case design studies, researchers must consider the ethical implications of removing treatment and withholding treatment during baseline. Withdrawal of treatment does not apply to this study, as the research team did not use a reversal design. The baseline condition in this study corresponded to business as usual, and because the treatment conditions constitute as over and beyond what is typically provided in classrooms, it was not unethical to withhold this additional treatment during the baseline.

Chapter IV: Results

This section reviews the preliminary analyses, including descriptive statistics and missing data, followed by the results of the analyses conducted to answer each research question. Interrater reliability and implementation integrity are reviewed, followed by visual analysis across participants for each of the three variables according to procedural standards put forth by What Works Clearinghouse (WWC). A non-overlap index, the Tau-*U* statistic, provides an effect size within the visual analyses to determine change in level between phases within participants. A summary of visual analysis findings is reported for each variable for overall treatment effect, differences in effect between training protocols, and differences in effect between conditions. Finally, a two-level multilevel model was run for each of the three variables: Overall Agreement, Disruptive Behavior Agreement, and Consequence Agreement.

The direct behavior rating form, the ISP-Function, is divided into two parts with five questions overall. The first part of the form has one question asking participants to rate the percentage of disruptive behavior that occurred in the observation on a scale of 0-10. The second part of the form has four questions that ask teachers to indicate the rate with which students' disruptive behavior is met with specific consequences (Adult Attention, Peer Attention, Escape/Avoidance, and Access to Tangibles or Activities). Visual analyses were conducted initially for the Overall Agreement (OA) using a difference score that combines the discrepancy across all five questions. Difference scores have been used in past research to evaluate rating accuracy (e.g., Kilgus et al., 2019). Preliminary visual analyses of Overall Agreement (OA) did not show a change in agreement levels following any of the trainings. Therefore, the research team conducted additional analyses to determine if agreement levels varied by type of rating. In

addition to the Overall Agreement, the researcher conducted visual analyses, Tau-*U* effect sizes, and a multilevel model for Disruptive Behavior Agreement (BA) and Consequence Agreement (CA). Disruptive Behavior Agreement used the discrepancy from the true score on only the first question on the ISP-Function. The Consequence Agreement used the combined discrepancy from the true score from the four consequence questions on the ISP-Function.

Incomplete Data

A high number of 'incomplete observations' may have influenced the results of the present investigation. In the visual analysis graphs, these are represented by open data points. When no disruptive behavior was observed in the observation period, the observation is considered incomplete. To illustrate, in a valid observation, the student may have five instances of disruptive behavior that spans across 20% of the observation according to the Systematic Direct Observation. If three of these five instances of disruptive behavior are met with adult attention, and one with peer attention, the conditional probability of adult attention and peer attention for this observation would be 60% and 20%, respectively. The teacher may rate the student's percentage of disruptive behavior as a three (30%), adult attention as a two (20%), and peer attention a zero. In this scenario, the disagreement would result in a difference score of 7 (one-point discrepancy for disruptive behavior, four for adult attention, and two for peer attention). During one incomplete observation, the student was playing quietly and interacting appropriately throughout the observation, and no disruptive behavior was observed. Alternatively, in an incomplete observation, the researcher recorded zero intervals of disruptive behavior, and subsequently, no consequences to disruptive behavior were recorded. Similarly, because the teacher also did not observe disruptive behavior, the teacher also rated all questions zero. This results in an agreement score of 100%.

Incomplete data were not removed due to the high frequency of these observations and the lack of remaining data if removed. In the absence of observed behavior, all ratings are zero, and agreement is typically perfect, or near-perfect with minimal variation, as aforementioned. Consequently, in analyzing agreement rates, these observations may result in skewed data and thus inaccurate conclusions drawn from such data. Below, the data are analyzed (1) in totality and (2) when incomplete data are removed. To account for the relationship between low rates of disruptive behavior on accuracy scores, the multilevel model used the percentage of disruptive behavior as a covariate to control for the effects of low rates of disruptive behaviors. From a total of 97 observations, 42 were considered incomplete, with zero instances of disruptive behavior. Across participants, 31 of 42 Baseline observations, 13 of 30 observations in Phase 1, and 11 of 25 observations in Phase 2 were considered valid, indicating that disruptive behavior was observed and rated within these observation periods. Participant 5's Baseline is noteworthy, given that two of Participant 5's incomplete data points had high levels of disagreement. Participant 5 rated the student's behavior as zero but rated adult and peer attention as the student received it naturally (not as a consequence of disruptive behavior). This is the only instance where incomplete data did not result in high levels of agreement and demonstrates a general lack of the teacher's understanding of the tool.

Interrater Reliability

To ensure interrater reliability of the observational data points, the research team participated in observation calibration and conducted interrater observations throughout the study. Reliability of at least 90% was considered acceptable. The agreement was calculated through a point to point ratio using the following equation, where A is the number of agreements, and D is the number of disagreements:

$$IOA = 100\,[\frac{A}{A + D}]$$

What Works Clearinghouse standards (Kratochwill et al., 2010) suggests agreement should be collected for at least 20% of all observational data points. Interobserver agreement was conducted for approximately 12% of observations due to barriers in available personnel and scheduling factors. The research team anticipated this barrier and therefore, all observers participated in a training and rater calibration to ensure reliability before the start of the study. A description of the calibration methods is described in Chapter 3. All agreement met acceptable levels of reliability (90%). The lowest agreement percentage was 96%, and the highest was 100%, with an average of 99.5%. If 100% agreement was not met initially, the discrepancies were discussed and resolved to record the observation's final score.

Implementation Fidelity

The researcher used an implementation protocol during each training with the participants that included a fidelity self-check list to ensure all components were reviewed. The fidelity self-checklists can be found in Appendix D. The fidelity checklists were created by the research team when creating the training protocol. The checklists included all critical elements of each training to ensure that trainings were implemented as designed. All fidelity checklists were reviewed at the end of the training by the researcher conducting the training to ensure each component had been completed within the training. If any training component was missing when completing the checklist, the missing component was completed before the training concluded. All trainings met 100% fidelity.

Visual Analysis of Overall Agreement

Figure 1 depicts the six graphs depicting each participants' observations across phases. This set of graphs represents the overall differences in agreement. For example, a difference score of 10 would indicate the participant's rating of the student was 10 points off the true score across all five questions. The highest possible difference score is a score of 50. In most cases, the

stability of Baseline, patterns within phases, and changes in level were skewed by incomplete observations where no disruptive behavior occurred. Therefore, the remainder of this visual analysis will look at patterns among valid data points only, as depicted by solid data points.

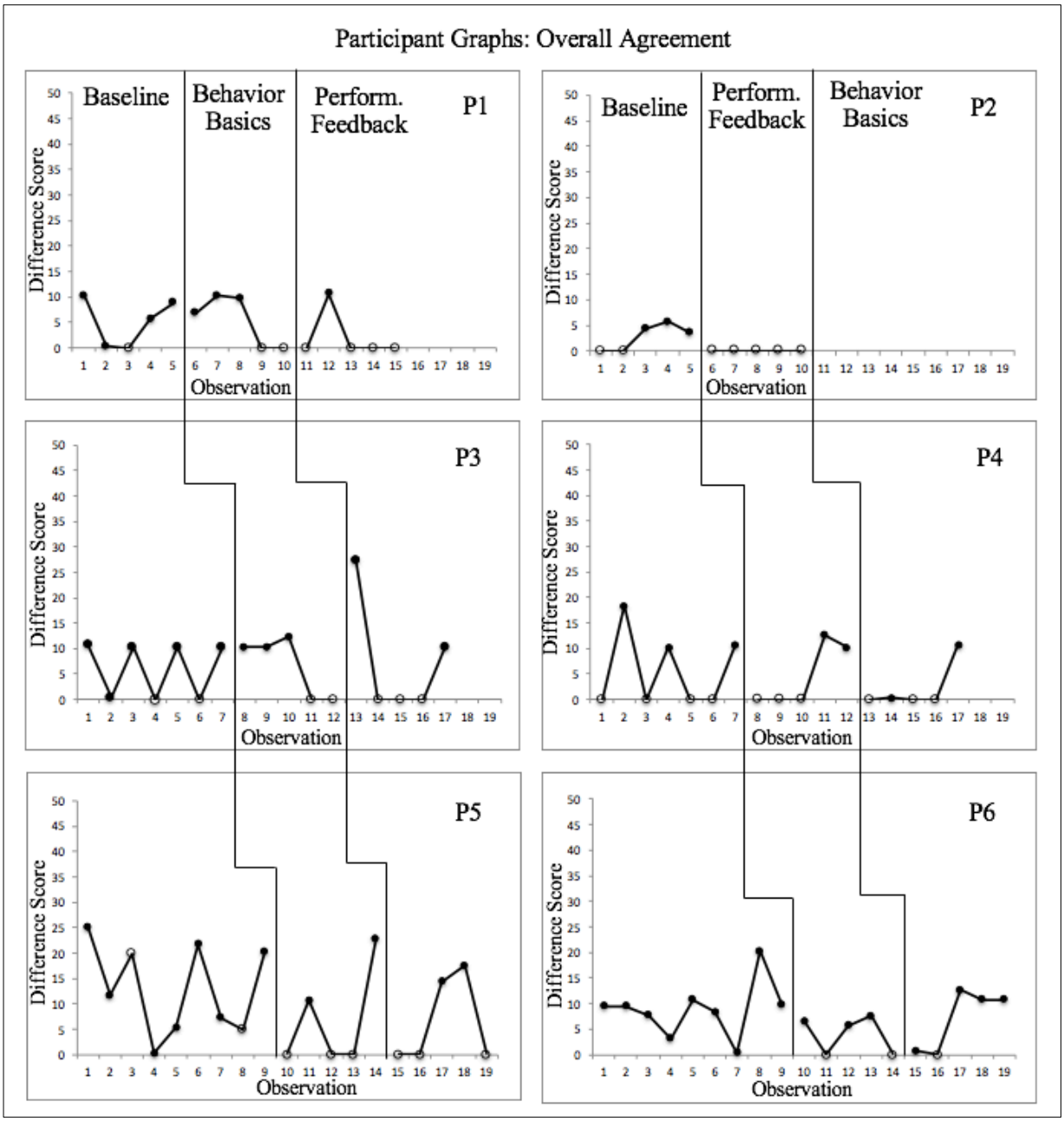


Figure 1. Participant graphs of Overall Agreement (OA) with phase change lines

Stability of Baseline. Participant 2 showed a stable Baseline with all three valid observations between seven and ten. Participants 1, 3, and 4 showed stability in Baseline with

one outlier. Participant 1 had agreement scores between 5.7 and 10.3, with one outlier at near-perfect agreement (0.3). Four out of the five valid data points for Participant 3 fell between 10.3 and 10.7, with one outlier observation showing near-perfect agreement. Two of Participant 4's three valid data points were within one (10.3 and 10.7) with one outlier at 18.4. Participant 5 showed the most instability with valid data points ranging from 0.3 to 25.1. Participant 6 showed some stability with a mean of 8.8 and six of nine data points falling between 7.7 and 10.7 and notably had no incomplete data points in Baseline. None of the participants' Baseline data showed a significant trend.

Phase 1, Condition 1. Participants 1, 3, and 5 were in Condition 1 and received the Behavior Basics training prior to all Phase 1 observations. Each participant had 2-3 incomplete data points within Phase 1. In looking at only valid data points, there does not appear to be a trend in Phase 1 for Participants 1 or 3. Participant 5 shows an upward trend towards disagreement within this phase. Participants 1 and 3 showed little variability, with all their valid observations in Phase 1 falling within a 4-point range. Participant 5 showed the most variability within Phase 1, with the two valid data points 12 points apart. Based on visual analyses, there was no noticeable change in Baseline level to Phase 1 when evaluating valid data points in any participant. Therefore, the immediacy of effect could not be evaluated. The Tau-*U* statistics seen below can range from -1 to +1. A Tau-*U* of -1 would indicate no overlap between pairs in the desired direction (negative). A score of +1.0 would indicate no overlap between pairs in the opposite direction, meaning the participants had higher disagreement across phases. Participants 1, 3, and 5 showed similar effect sizes at 0.34, 0.34, and 0.29, respectively, all trending towards higher disagreement scores from Baseline to Phase 1.

Phase 1, Condition 2. Participants 2, 4, and 6 were in Condition 2 and received the Performance Feedback training in Phase 1. As aforementioned, all participants had 2-3

incomplete data points, except Participant 2, who had all incomplete data points. Therefore, no patterns, trends, or variability could be observed in Participant 2. Participant 4 shows a slight downward trend towards agreement. There is not a trend in Phase 1 for Participant 6. Participants 4, and 6 showed little variability, with all their valid observations in Phase 1 falling within a 4-point range. No patterns, trends, or variability could be observed in Participant 2, as there were no valid data points within Phase 1. Based on visual analyses, there was not a noticeable change in level from Baseline to Phase 1 when evaluating valid data points in any participant. Therefore, the immediacy of effect could not be evaluated. The Tau-*U* for Participants 4 and 6 in Phase 1 both indicated improvements in accuracy scores. Participant 6 saw the largest effect size of -0.56, and Participant 4 also saw a favorable effect size of -0.17.

Phase 2, Condition 1. Participants 1, 3, and 5 received the Performance Feedback training in Phase 2. Phase 2 had more incomplete data points than Phase 1 across participants. Participant 1 only had one valid data point, so no trends or patterns within this phase could be analyzed. Participants 3 and 5 each had two valid data points within Phase 2. Participant 3 showed a downward trend towards agreement, while Participant 5 showed an upward trend towards disagreement within the valid data points in Phase 2. Generally, across participants, more variability was shown within participants in Phase 2. Participants 3 showed the most Phase 2 variability with the two valid data points varying by 17 points, and Participant 5 showed less variability with a 3.1-point range. There was no noticeable change in level from Phase 1 to Phase 2 when evaluating valid data points in any participant in Condition 1, and therefore, the immediacy of effect could not be evaluated. The Tau-*U* statistic supports the visual analysis indicting a lack of treatment effect with effect sizes ranging from 0.0 to +1.0 (see table 3). Participants 1 and 3 both saw the undesirable effect size of 1.0 and 0.34 trending towards higher

disagreement, respectively. Finally, Participant 5 showed substantial overlap with a Tau-*U* of 0, indicating no changes in level from Phase 1 to Phase 2.

Phase 2, Condition 2. Condition 2 participants (Participant 4 and 6) received the Behavior Basics training prior to Phase 2 observations. Participant 2 withdrew from the study and therefore did not complete Phase 2. As aforementioned, Phase 2 had more incomplete data points than Phase 1 across participants. Participant 4 had two valid data points within Phase 2 and showed an upward trend towards disagreement. Participants 4 showed high variability in Phase 2, with the two valid data points varying 10.4 points. Participant 6 showed less variability with the first observation, approximately 11 points lower than the final three data points, which fell within a 2-point range. Participant 6 showed stability within the last three data points in the phase with one outlier. Based on visual analyses, there was no noticeable change in level from Phase 1 to Phase 2 in either Condition 2 participant. The immediacy of effect could not be evaluated. The lack of change in level is supported by the Tau-*U* statistic, which was calculated using valid data points only (see table 3). The Tau-*U* yielded a wide range of effect sizes between -0.50 to +0.5 from Phase 1 to Phase 2. Participant 4 saw the largest desirable effect size with a Tau-*U* of -0.50, indicating that pairs trended towards agreement. Conversely, Participant 6 showed positive effect sizes trending towards disagreement with Tau-*U* effect sizes of 0.5.

Table 3: Tau-U Statistics with Valid Data Only for Overall Agreement (OA)

Participant Number	Ba-P1	P1-P2
Participant 1	0.34	1.00
Participant 3	0.34	0.34
Participant 4	-0.17	-0.50
Participant 5	0.29	0.00
Participant 6	-0.56	0.50

Note: Ba=Baseline; P1=Phase 1; P2=Phase 2
Negative Tau-U scores indicate scores trending towards agreement between phases.

Summary of Visual Analysis for Overall Agreement. In evaluating valid data points, Participants 1, 2, 3, and 4 had stable Baselines with little variability. Participants 5 and 6 had

higher variability and little stability in the Baseline phase. None of the participants showed trends in Baseline except Participant 4, who showed a slight downward trend. Based on the visual analysis, there was no visually detectable difference in level in any of the Phases across participants. Given the lack of observable treatment effects, no differences between the effect of Behavior Basics and Performance Feedback training can be evaluated or the effect of the order of trainings based on the visual analyses. Table 4 shows the Tau-*U* effect sizes for Overall Agreement (OA) by participant (note these are the same values as in Table 3, just organized by treatment as opposed to phase). Ultimately across participants, there was wide variability in Tau-*U* scores ranging between -0.50 and 1.00. Based on the Tau-*U* effect sizes, two participants (Participants 4 and 6) saw decreases in difference scores following the Performance Feedback training. In contrast, only one saw a decrease in difference scores following the Behavior Basics training (Participants 4). Participant 4 showed improved agreement following both treatments based on the Tau-*U* effect sizes.

Table 4: Tau-U Statistics by Treatment for Overall Agreement (OA)

Participant Number	Behavior Basics	Performance Feedback
Participant 1	0.34	1.00
Participant 3	0.34	0.34
Participant 4	-0.5	-0.17
Participant 5	0.29	0.00
Participant 6	0.5	-0.56

Note: Participants 1, 3, and 5 received Behavior Basics first. Participants 4 and 6 received Performance Feedback first.
Negative Tau-U scores indicate scores trending towards agreement between phases.

Visual Analysis of Disruptive Behavior Agreement (BA)

Figure 2 depicts the six graphs depicting each participant's observations across phases for their agreement in rating disruptive behavior only. For example, a difference score of 5 would indicate the participant's rating of the student was 5 points off the true score in only the disruptive behavior rating. The highest possible difference score is a score of 10. In most cases, the stability of Baseline, patterns within phases, and changes in level were skewed by incomplete

observations where no disruptive behavior occurred. Therefore, the remainder of this visual analysis will look at patterns among valid data points only, as depicted in the graph by solid data points.

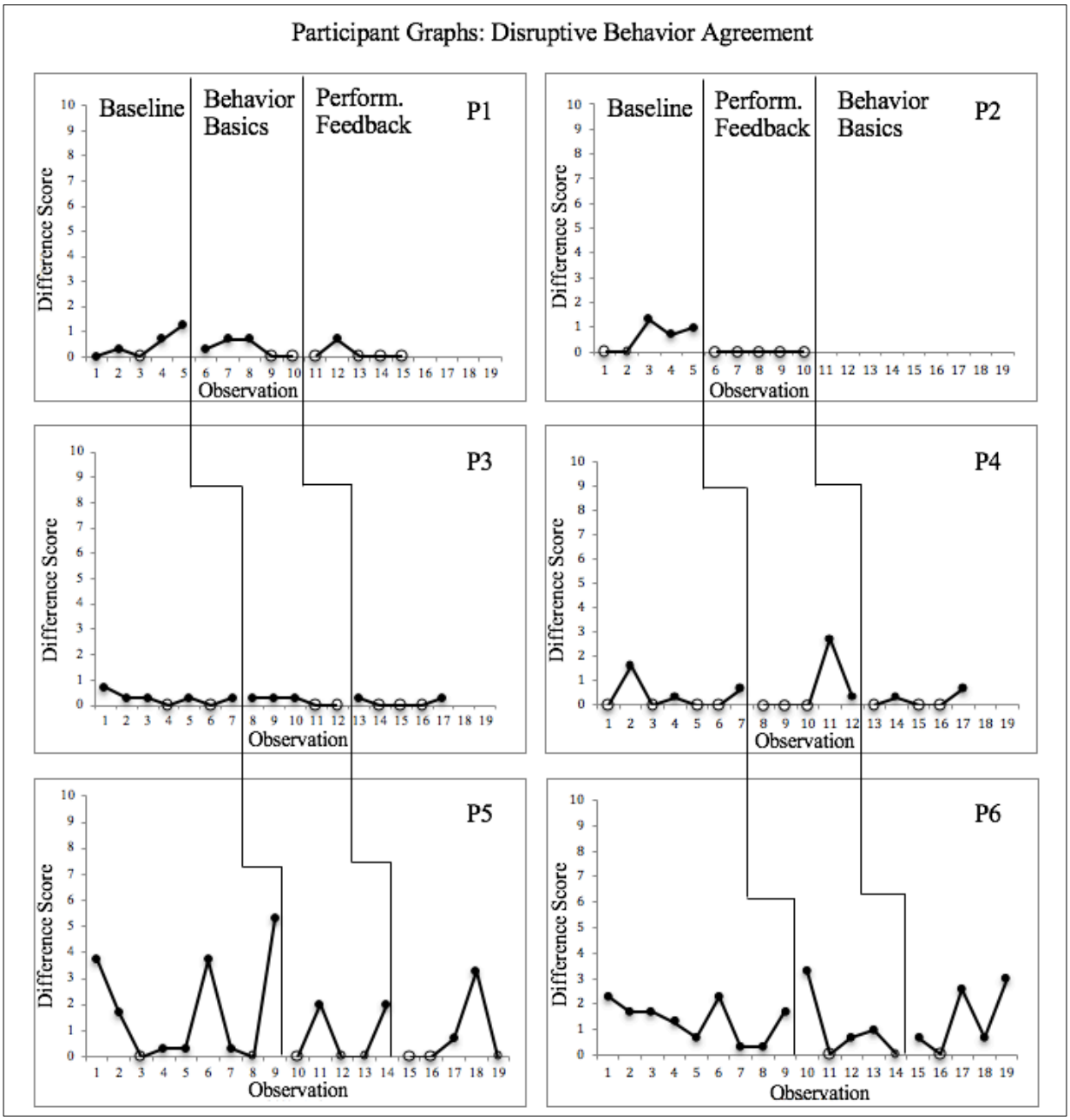


Figure 2. Participant graphs of Disruptive Behavior Agreement (BA) with phase change lines

Stability of Baseline. A stable Baseline was observed in Participants 1, 2, 3, and 4, and all had little variability between data points with no more than a 1.3-point range between the highest and lowest agreement score. Participant 6 did not show stability in Baseline and had a slightly larger variability among Baseline data points with a range of 2 points between the lowest and highest disagreement scores. Participant 5 had the most variability with disagreement scores ranging from 0.3 to 5.3 and did not show stability in Baseline. Participant 1 showed a slight trend upward towards disagreement, and Participants 4 and 6 showed a slight downward trend towards agreement. Participants 2, 3, and 5 did not have any trends in their Baseline observations. Notably, all participants except Participant 6 had low disagreement scores, with all scores under 2.3. Participant 6 was the exception with three observations with scores higher than 3.

Phase 1, Condition 1. Condition 1 participants received the Behavior Basics training following baseline, prior to Phase 1 observations. Participants 1 and 3 showed low scores in Phase 1 with all scores under one, and all participants had all scores under 3.5. Participants 1, 3, and 5 did not show a trend within Phase 1 and showed little variability with all observations within 0.5 points. The visual analysis did not show a change in Baseline level to Phase 1 in any of the participants. However, floor effects may have influenced this as there were very low levels of disagreement in Baseline across participants. The Tau-*U* statistic supports this conclusion, which was calculated using only the valid observations (see table 5). The Tau-*U* yielded effect sizes between -0.20 to +0.14 from Baseline to Phase 1. Participant 3 saw the largest desirable effect size with a Tau-*U* of -0.20, indicating that pairs trended towards agreement. Participants 1 and 5 both showed small positive Tau-*U*'s of 0.08 and 0.14, respectively, indicating minimal changes in level from Baseline to Phase 1 with pairs trending towards disagreement.

Phase 1, Condition 2. Condition 2 participants (Participants 2, 4, and 6) received the Performance Feedback training following baseline. However, Participant 2's data in Phase 1

could not be interpreted given there were no valid data points in Phase. Within Phase 1, Participants 4 and 6 show a downward trend towards agreement and some variability within Phase 1 with a 2.4-point range for Participant 4 and a 2.6-point range for Participant 6. The visual analysis did not show a noticeable change in Baseline level to Phase 1 in any of the participants when evaluating valid data points. Low levels of disagreement in Baseline likely influenced the limited change in level due to floor effects. The Tau-*U* statistic supports this conclusion with both Participant 4 and Participant 6 having small effect sizes of 0.17 and 0.07, respectively, trending towards higher disagreement.

Phase 2, Condition 1. Condition 1 participants (Participants 1, 3, and 5) received the Performance Feedback training before Phase 2 observations. Like Phase 1, Participant 1 and 3 showed consistently low data points across valid observations with all difference scores under 1 in Phase 2. Participant 1 only had one valid data point, and therefore, trends or variability could not be evaluated for Participant 1. Participant 3 showed high stability of data points with no trend. Participants 5 showed more variability among data points with scores ranging from 0.7-3.3 and a slight upward trend towards disagreement. There was no observable decrease in level from Phase 1 to Phase 2 in any of the Condition 1 participants, which is confirmed by the Tau-*U* effect sizes (see Table 5). Participant 1 saw the most undesirable effect size of +0.33. Participants 3 and 5 both showed no change in level with Tau-*U* effect sizes of 0.00.

Phase 2, Condition 2. Participants 4 and 6 received the Behavior Basics training in Phase 2. No data was collected for Participant 2 in Phase 2 due to attrition. None of the Condition 2 participants showed a trend within Phase 2. Participant 4 showed consistently low data points across valid observations with all difference scores under 1.0 and little variability, with no more than a 0.5-point difference between the lowest and highest data point in Phase 2. Participant 6 showed more variability among data points, with scores ranging from 0.7-3.0.

Based on visual analyses, only Participant 4 showed a decrease in difference scores between phases. The Tau-*U* effect sizes calculated support the visual analyses (see table 5). Participant 4 saw the largest desirable effect size with a Tau-*U* of -0.25. Participant 6 showed a weaker but desirable effect size with a Tau-*U* of -0.17, indicating a small change in level towards agreement.

Table 5: Tau-U Statistics with Valid Data Only for Disruptive Behavior Agreement (BA)

Participant Number	Ba-P1	P1-P2
Participant 1	0.08	0.33
Participant 3	-0.20	0.00
Participant 4	0.17	-0.25
Participant 5	0.14	0.00
Participant 6	0.07	-0.17

Note: Ba=Baseline; P1=Phase 1; P2=Phase 2
Negative Tau-U scores indicate scores trending towards agreement between phases.

Summary of Visual Analysis of Disruptive Behavior Agreement. Participants 1, 2, 3, and 4 showed stability in Baseline, and Participants 5 and 6 showed higher amounts of variability within the Baseline phase. Notably, all participants except Participant 5 had low disagreement scores, with all scores under 2.3. Based on visual analyses, only Participant 4 showed a decrease in difference scores between Phase 1 and Phase 2. No other changes in level were observed in any participant between any phases. Given the lack of observable treatment effects, no differences between the effect of Behavior Basics training and the Performance Feedback training or the effect of the order of trainings can be evaluated based on the visual analysis. Table 6 shows the Tau-*U* effect sizes for Disruptive Behavior Agreement (BA) by participant (note these are the same values as in Table 5, just organized by treatment as opposed to phase). Ultimately, Tau-*U* scores across participants ranged between -0.25 and 0.17, indicating very small effect sizes and little change in agreement scores between phases. Based on the Tau-*U* effect sizes, three participants (Participants 3, 4, and 6) saw decreases in difference scores

following the Behavior Basics training. In contrast, none of the participants saw decreases in difference scores following the Performance Feedback training.

Table 6: Tau-U Statistics by Treatment for Disruptive Behavior Agreement (BA)

Participant Number	Behavior Basics	Performance Feedback
Participant 1	0.08	0.33
Participant 3	-0.20	0.00
Participant 4	-0.25	0.17
Participant 5	0.14	0.00
Participant 6	-0.17	0.07

Note: Participants 1, 3, and 5 received Behavior Basics first. Participants 4 and 6 received Performance Feedback first.
Negative Tau-U scores indicate scores trending towards agreement between phases.

Visual Analysis of Agreement for Rating Consequences

Figure 3 depicts the six graphs depicting each participant's observations across phases for their agreement in rating the consequences of the student's behavior only. For example, a difference score of 15 would indicate the participant's rating of the student was 15 points off the true score across the four consequence ratings only. The highest possible difference score is a score of 40. In most cases, the stability of Baseline, patterns within phases, and changes in level were skewed by incomplete observations where no disruptive behavior occurred. Therefore, the remainder of this visual analysis will look at patterns among valid data points only, which is depicted in the graph by solid data points.

Stability of Baseline. Participants 1, 2, 3, and 4 saw stable Baselines with little variability, although Participants 1, 3, and 4 had one outlier data point. Participant 2 showed the least amount of variability with all difference scores between 2.6 and 5.0. Participant 5 showed the most instability and variability within Baseline with difference scores ranging from 0.0-21.4. Participant 6 also did not show stability in Baseline and had high variability between observations with difference scores ranging from 0.0-20.0. Participant 4 showed a slight trend downward towards agreement, and no other participants showed any trends within Baseline.

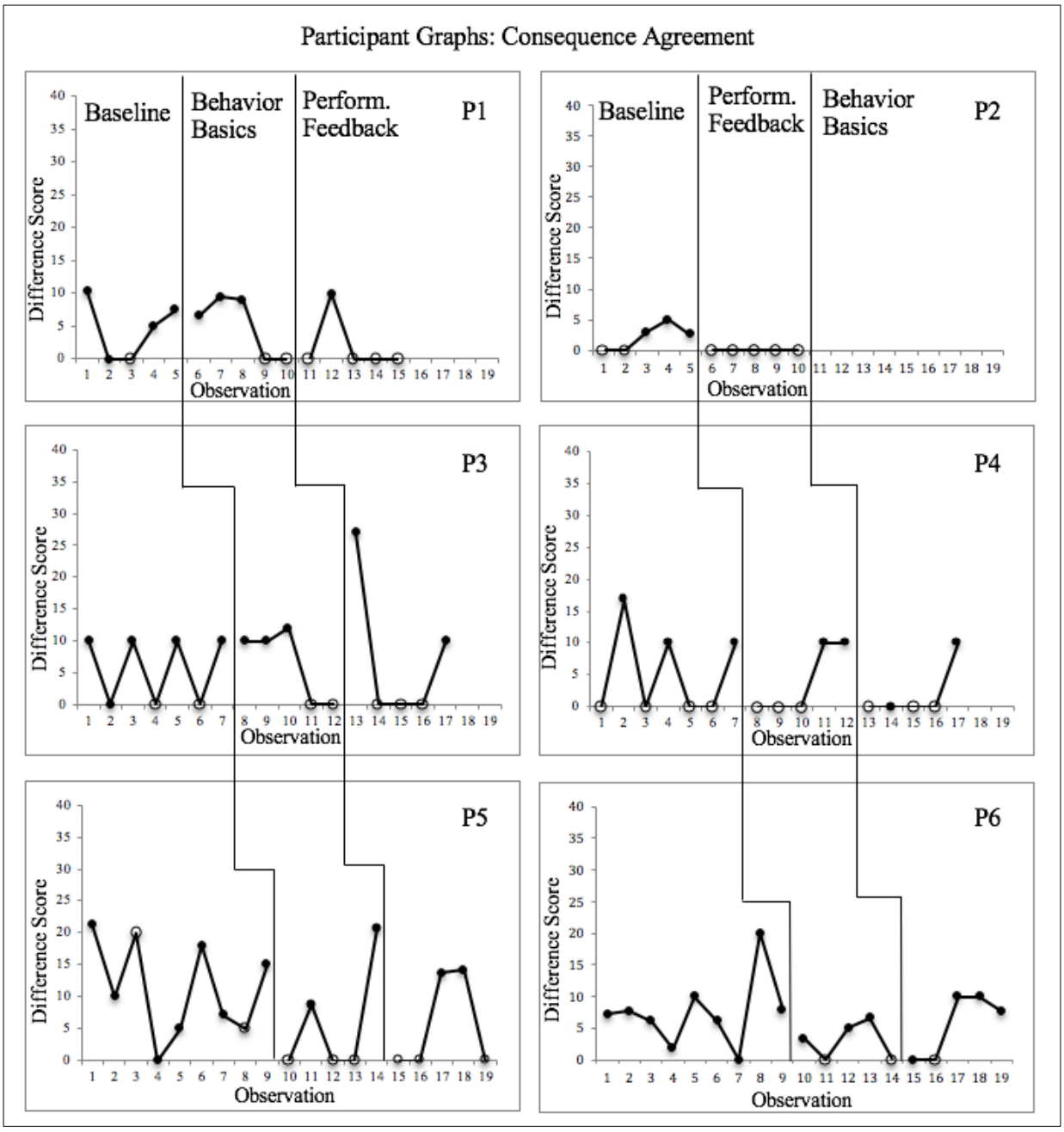


Figure 3. Participant graphs of Consequence Agreement (CA) with phase change lines.

Phase 1, Condition 1. Participants 1, 3, and 5 received the Behavior Basics training prior to Phase 1 observations. Data within Phase 1 for Participants 1 and 3 showed stability with no clear trends and little variability with point ranges of 3.0 and 2.0, respectively. Participant 5 showed an upward trend towards disagreement within valid observations and showed high

variability between observations within this phase with a 12-point difference between the lowest and highest data point. None of the participants in Condition 1 showed a noticeable decrease in level from Baseline to Phase 1. The immediacy of effect could not be evaluated within the visual analysis due to a lack of noticeable change in level. This conclusion is supported by the Tau-*U* statistic, which was calculated using valid observations (see table 7). Participants 1, 3, and 5 showed positive effect sizes trending towards disagreement with Tau-*U*'s of 0.33 and 0.29, respectively. Participant 3 saw the most undesirable effect size of 0.47 trending towards a higher difference score.

Phase 1, Condition 2. Participants 2, 4, and 6 received the Performance Feedback directly following baseline, prior to Phase 1 observations. No patterns, trends, or variability could be observed in Participant 2, as there were no valid data points within Phase 1. Participant 4 showed stability with no clear trends and little variability, with both valid data points falling at 10.0. Participant 6 showed a slight upward trend towards disagreement and more variability with a 3.3-point difference between the lowest and highest data point. Based on visual analyses for consequence agreement, only Participant 6 showed a small change in level, given the increased stability of lower difference scores in Phase 1 from Baseline. However, the Tau-*U* statistics showed both Condition 2 participants had small but desirable changes in level. Participant 4 and 6 had desirable effect sizes with a Tau-*U* of -0.33 and -0.41, respectively.

Phase 2, Condition 1. Participants 1, 3, and 5 received the Performance Feedback training after Phase 1. Participant 1 only had one data point, and therefore trends and variability between data points could not be analyzed. Participant 3 showed a downward trend towards agreement and high variability among Phase 2 observations with a difference score range of 17.0 points. Participant 5 did not show any trends or variability with only a 0.5-point difference between observations. None of the participants appeared to have a significant change in level

from Phase 1 to Phase 2 when evaluating the accuracy of consequence ratings. The Tau-*U* statistic, which was calculated using valid data points only (see table 7), yielded a range of effect sizes between 0.0-1.0. Participant 1 saw the most undesirable effect size of 1.0 trending towards higher disagreement. Participant 3 also showed a positive effect size trending towards disagreement with a Tau-*U* of 0.33. Finally, Participant 5 showed no changes in level from Phase 1 to Phase 2 with a Tau-*U* effect size of 0.00

Table 7: Tau-U Statistics with Valid Data Only for Consequence Agreement (CA)

Participant Number	Ba-P1	P1-P2
Participant 1	0.33	1.00
Participant 3	0.47	0.33
Participant 4	-0.33	-0.50
Participant 5	0.29	0.00
Participant 6	-0.41	0.50

Note: Ba=Baseline; P1=Phase 1; P2=Phase 2
Negative Tau-U scores indicate scores trending towards agreement between phases.

Phase 2, Condition 2. Participants 4 and 6 received the Behavior Basics training second, following Phase 1. Participant 2 dropped out of the study after Phase 1 and therefore had no data points in Phase 2. Participant 4 showed an upward trend towards disagreement, and Participant 6 did not show any trends within this phase. Both Participants 4 and 6 showed high variability among Phase 2 observations with difference score ranges of 10.0 points, each. Participant 4 had two valid data points within this phase. The first observation showed high agreement and an improvement from Phase 1, and the second observation showed no change in agreement from Phase 2. Given the discrepancy between these scores, a sustained change in level was not observed. Participant 6 appeared to have an undesirable change in level or increase in difference scores from Phase 1 to Phase 2 in rating consequences based on visual analysis. The Tau-*U* statistic, which was calculated using valid data points only (see table 7), supports the visual analysis that there was a decrease in level for one of the data points for Participant 4 and an increase in level in Participant 6. Participant 4 saw the largest desirable effect size with a Tau-*U*

of -0.50, indicating pairs trended towards agreement. Participant 6 showed positive effect sizes trending towards disagreement with a Tau-*U* of 0.5.

Summary of Visual Analysis of Consequence Agreement. Participants 1, 2, 3, and 4 saw stable Baselines with little variability within Baseline. Participants 5 and 6 did not have stable Baselines for rating consequences, and Participant 5 showed the most instability and variability. Participant 4 showed a slight trend downward towards agreement, and no other participants showed any trends within Baseline. Based on visual analyses, only Participant 6 showed small changes in level. Participant 6 appeared to show improved scores from Baseline to Phase 1 and high disagreement scores in Phase 2. Participant 4 also showed improvement from Phase 1 to Phase 2 in one of two observations. No other participants showed changes in level between phases based on the visual analysis. Given the lack of observable treatment effects, no differences between the effect of Behavior Basics training and the Performance Feedback training. Further, the lack of observable treatment effects prevents conclusions from being drawn regarding the effect of the order of trainings based on visual analysis. Table 8 shows the Tau-*U* effect sizes for Consequence Agreement (CA) by participant (note these are the same values as in Table 7, just organized by treatment as opposed to phase). Ultimately across participants, Tau-*U* scores ranged between -0.50 and 1.00. Based on the Tau-*U* effect sizes, two participants (Participants 4 and 6) saw decreases in difference scores following the Performance Feedback training. In contrast, only one of the participants (Participant 4) saw decreases in difference scores following the Behavior Basics training. Only Participant 4 saw decreases following both trainings.

Table 8: Tau-U Statistics by Treatment for Consequence Agreement (CA)

Participant Number	Behavior Basics	Performance Feedback
Participant 1	0.33	1.00
Participant 3	0.47	0.33
Participant 4	-0.50	-0.33
Participant 5	0.29	0.00
Participant 6	0.50	-0.41

Note: Participants 1, 3, and 5 received Behavior Basics first. Participants 4 and 6 received Performance Feedback first.
Negative Tau-U scores indicate scores trending towards agreement between phases.

Multilevel Model

Hierarchical Linear Modeling (HLM) was utilized to determine average treatment effects across and within participants. A two-level model was used to analyze the data with individual observations nested within participants to estimate the average change in level across phases, the variance in Baseline levels, and the variance in treatment levels for each variable. Additionally, the percentage of disruptive behavior was built into the model as a covariate to control for the incomplete observations. Differences in phase levels were compared during Baseline to Phase 1 and Phase 2. The following model was applied for all participants and then separately for each condition to determine differences in overall treatment effects and differences in effects based on the order the trainings were administered.

Level-1 Equation

$$Y_{ij} = \beta_{0j} + \beta_{1j}\text{cov} + \beta_{2j}\text{BB} + \beta_{3j}\text{BB*cov} + \beta_{4j}\text{BB*condition} + \beta_{5j}\text{BB*condition*cov} + \beta_{6j}\text{PF} + \beta_{7j}\text{PF*cov} + \beta_{8j}\text{PF*condition} + \beta_{9j}\text{PF*condition*cov} + e_{ij}$$

Level-2 Equations

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + u_{2j}$$

where Y_{ij} is the agreement score at time i for case j, cov is the percentage of disruptive behavior in each observation, serving as a covariate, and scaled such that a value of 0 corresponds to no disruptive behavior, BB is dummy coded with values of 1 for observations from a Behavior Basics phase and 0 for all other observations, PF is dummy coded with values of 1 for observations from a Performance Feedback phase and 0 for all other observations, and condition is dummy coded with values of 0 for observations that came from cases where the treatment phase order was BB followed by PF and values of 1 for observations that came from cases where the order of treatment phase was PF followed by BB., As a consequence of this coding, β_{0j} is the predicted level of agreement in Baseline when the value of cov is 0, β_{1j} is the effect of the percent of disruptive behavior (i.e., the covariate) on the predicted level of agreement during baseline, β_{2j} is the difference in the expected level of agreement when cov=0 for observations in a Behavior Basics phase and observations in baseline when the Behavior Basics phase immediately follows baseline, β_{3j} indicates the amount this expected difference changes with each unit change in the covariate, β_{4j} indexes the amount the covariate adjusted the expected difference between Behavior Basics and baseline phases changes when Behavior Basics is the second treatment phase, as opposed to the first treatment phase, β_{5j} indexes the amount this change varies with the covariate, β_{6j} is the difference in the expected level of agreement when cov=0 for observations in a Performance Feedback phase and observations in baseline when the Performance Feedback phase is the second treatment phase, β_{7j} indicates the amount this expected difference changes with each unit change in the covariate, β_{8j} indexes the amount the covariate adjusted expected difference between Performance Feedback and baseline phases changes when Performance Feedback is the first treatment phase, as opposed to the second treatment phase, and β_{9j} indexes the amount this change varies with the covariate.

Table 9: Results of Multilevel Models by Variable

Model	Parameter	Parameter Estimate	Standard Error	CI Lower	CI Upper	*P*
Overall Agreement	*Fixed Effects*					
	Intercept	4.57	1.38	1.71	7.44	0.003**
	Covariate (cov)	0.35	0.12	0.11	0.59	0.005**
	Basics	-1.21	2.18	-5.54	3.13	0.582
	Basics*cov	-0.09	0.14	-0.37	0.19	0.544
	Basics*condition	-2.04	3.26	-8.6	4.52	0.535
	Basics*condition *cov	0.15	0.2	-0.25	0.55	0.456
	Performance	-1.61	2.22	-6.65	3.43	0.488
	Perform*cov	-0.16	0.14	-0.43	0.11	0.241
	Perform*condition	-0.66	2.76	-8.06	6.74	0.822
	Perform*condition*cov	0.04	0.22	-0.4	0.48	0.849
	Variance Estimates					
	Intercept	1.15	2.83	--	--	0.342
	Basics	0	--	--	--	--
	Performance	0.47	6.42	--	--	0.471
	Residual	32.92	5.22	--	--	<0.0001***
Disruptive Behavior Agreement	*Fixed Effects*					
	Intercept	0.2	0.19	-0.18	0.58	0.3
	Covariate (cov)	0.08	0.02	0.05	0.11	<0.0001***
	Basics	0.003	0.28	-0.56	0.57	0.99
	Basics*cov	-0.06	0.02	-0.1	-0.03	0.0011**
	Basics*condition	-0.19	0.42	-1.03	0.65	0.654
	Basics*condition *cov	0.08	0.03	0.03	0.13	0.0025**
	Performance	-0.04	0.28	-0.61	0.52	0.876
	Perform*cov	-0.07	0.02	-0.1	-0.03	0.0002**
	Perform*condition	0.04	0.35	-0.66	0.74	0.909
	Perform*condition*cov	0.09	0.03	0.03	0.14	0.0014**
	Variance Estimates					
	Intercept	0.04	0.05	--	--	0.22
	Basics	0	--	--	--	--
	Performance	0	--	--	--	--
	Residual	0.57	0.09	--	--	<0.0001***

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

Table 9 Continued: Results of Multilevel Models by Variable

Model	Parameter	Parameter Estimate	Standard Error	CI Lower	CI Upper	*P*
Conse-quence Agreement	*Fixed Effects*					
	Intercept	4.4	1.29	1.71	7.09	0.003**
	Covariate (cov)	0.27	0.11	0.04	0.5	0.023*
	Basics	-1.19	2.05	-5.27	2.9	0.564
	Basics*cov	-0.02	0.13	-0.29	0.24	0.865
	Basics*condition	-1.88	3.08	-8.11	4.36	0.546
	Basics*condition *cov	0.07	0.19	-0.31	0.45	0.719
	Performance	-1.58	2.14	-6.52	3.36	0.483
	Perform*cov	-0.09	0.13	-0.35	0.17	0.496
	Perform*condition	-0.75	2.7	-8.02	6.53	0.795
	Perform*condition*cov	-0.05	0.21	-0.46	0.37	0.825
	Variance Estimates					
	Intercept	0.82	2.51	--	--	0.372
	Basics	0	--	--	--	--
	Performance	1.17	6.29	--	--	0.426
	Residual	29.1	4.61	--	--	<0.0001***

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

Each HLM was conducted assuming a change in levels between phases and no change in trend as preliminary visual analyses did not indicate any significant trends. The data was run using SAS Software, Version 9.4 (SAS Institute, 2015) with PROC MIXED.

Overall Agreement (OA). Results from the multilevel model for Overall Agreement (OA) indicated the intercept was statistically significant (p=.0003). Further, the covariate, percentage of disruptive behavior in the observation was statistically significant for OA ratings (p = 0.005), indicating that the covariate was significantly related to OA scores, with the positive coefficient indicating that the difference in ratings tended to be greater when the percentage of disruptive behavior was higher. The remaining main effects and interaction effects were not statistically significant, for difference scores of OA ratings, indicating that there was no

significant difference in OA scores following the Behavior Basics training or Performance Feedback training in either condition while controlling for the percentage of disruptive behavior.

Disruptive Behavior Agreement (BA). Results from the multilevel model indicated that the covariate was statistically significant for Disruptive Behavior Agreement (BA) ratings (p=<.0001), indicating the percentage of disruptive behavior within the observation was significantly related to BA scores, again with difference in ratings being greater when the percentage of disruptive behavior was higher. The interaction effects between the covariate and the Behavior Basics training in both Condition 1 and 2 were statistically significant (p=0.0011, p=0.0025, respectively) as well as the interaction effects between the covariate and Performance Feedback training in both Condition 1 and 2 (p=0.0002, p=0.0014, respectively). The coefficient for each of these interactions was positive, indicating the tendency for the difference in rating to be higher when there was more disruptive behavior was more pronounced in the treatment phases than the baseline phase. The main effects were not statistically significant for difference scores of BA ratings indicating that there was not a significant difference in BA scores following the Behavior Basics training or Performance Feedback Training in either condition while controlling for the percentage of disruptive behavior.

Consequence Agreement (CA). Results from the multilevel model indicated that the intercept for Consequence Agreement (CA) was statistically significant (p=.003). The covariate, percentage of disruptive behavior in the observation, was statistically significant for CA ratings (p = 0.023), indicating that the covariate was significantly related to CA scores. The positive coefficient indicates the difference in ratings tended to be larger when there was more disruptive behavior. The remaining main effects and interaction effects were not statistically significant for difference scores of CA ratings, indicating that there was not a significant difference in CA

scores following the Behavior Basics training or Performance Feedback training, in either condition while controlling for the percentage of disruptive behavior.

Summary of Results

The present study explored the changes in accuracy direct behavior ratings of preschool teacher ratings of disruptive behaviors following two training protocols. Preliminary visual analyses of Overall Agreement (OA) did not show a change in agreement levels following any of the trainings. Therefore, the primary researcher conducted additional analyses to determine if agreement levels varied by type of rating by conducting visual analyses, Tau-*U* effect sizes, and a multilevel model for each of the three variables: Overall Agreement, Disruptive Behavior Agreement, and Consequence Agreement. Consequence Agreement accounted for 78.5% of the disagreement in scores across all participants, across phases.

Based on the visual analysis for Overall Agreement, no participants showed a change in level of agreement following either trainings in either condition. Participants 1, 2, 3, and 4 showed stable Baselines, but ultimately many participants showed variability within Phases, making it difficult to draw conclusions about consistent changes in level. Tau-*U* effect sizes indicated decreases in Participant 4 and Participant 6 following the Performance Feedback trainings that fall in the small effect size range (-0.5, -0.56, respectively). Participant 4 also saw an improvement, but with a small effect size following the Behavior Basics training (-0.17). No other participants had effect sizes that indicated an improvement in accuracy scores for Overall Agreement. The results of the multilevel model support the conclusion that there were no statistically significant improvements in accuracy for Overall Agreement. While the intercept and covariate were statistically significant for Overall Agreement (p=.0003; p = 0.005, respectively), no other main effects or interaction effects were statistically significant, indicating there was not a significant change in level between phases for any training in any condition.

The results for Disruptive Behavior Agreement showed very low Baseline levels of disagreement in the visual analysis and the multilevel model with a non-significant intercept. All agreement scores across participants in Baseline were under a disagreement score of 2.3 out of 10. Based on the visual analysis, only Participant 4 showed a visually detectable change in level following the Behavior Basics training. Tau-*U* effect sizes were small across participants and phases. Participants 3, 4, and 6 showed desirable changes in level, but ultimately small effect sizes following the Behavior Basics training based on the Tau-*U* statistics (-0.20, -0.25, and -0.17, respectively). None of the participants saw decreases in difference scores following the Performance Feedback training based on Tau-*U* effect sizes. The multilevel model supports these findings with no statistically significant main effects. However, the multilevel model did reveal that the covariate, percentage of disruptive was statistically significant (p=<0.0001), and the interaction effects within the Behavior Basics and Performance Feedback phases in both conditions.

Finally, the Consequence Agreement's findings showed similar patterns as Overall Agreement as 78.5% of the disagreement in Overall Agreement discrepancies were based on the final four questions. Participants 1-4 showed stable Baselines, but many participants showed high variability within phases, making it difficult to determine a consistent change in level. Based on visual analyses alone, only Participant 6 showed any changes in level with improvement following the Performance Feedback training and showed an increase following the Behavior Basics training. The Tau-*U* statistic produced small effect sizes in the desirable direction for Participant 4 and 6 (-0.33, -0.41) following the Performance Feedback training and Participant 4 following the Behavior Basics training (-0.050). The multilevel model for the Consequence Agreement showed the intercept and covariate were statistically significant (p=.0003; p = 0.023, respectively), but no other main effects or interaction effects were

statistically significant indicating there was not a significant change in level between phases for any training in any condition. Together, these results are integrated in Chapter 5 to answer the three primary research questions. However, due to the lack of observable treatment effects across participants, phases, and the three variables evaluated, no differences between the effect of Behavior Basics training and the Performance Feedback training or the effect of the order of trainings can be evaluated to answer the final two research questions.

Chapter V: Discussion

This chapter discusses the interpretation of the above cumulative results, including the visual analysis, Tau-*U* effect sizes, and multilevel model. Additionally, this chapter considers the limitations encountered by the research team and its implications on the results and future redirections in research. Finally, the implications of these findings for practice are reviewed.

Response to Research Question 1

To answer the first research question, "To what extent does teacher training improve the accuracy of ISP-Function data collected?", the researcher conducted a visual analysis, Tau-U effect sizes, and reviewed the multilevel to determine if there was an effect of the trainings on the accuracy of teacher ratings of the ISP-Function. Based on the results from this study, the teacher trainings conducted did not improve the accuracy ISP-Function data collected. While several participants displayed small improvements in accuracy scores, improvements were largely inconsistent, all Tau-*U* effect sizes were in the small range, and no significant improvements were demonstrated across participants based on the visual analysis. Further, the multilevel model indicated no statistically significant main effects for Overall Agreement, Disruptive Behavior Agreement, or Consequence Agreement.

Response to Research Question 2

To respond to the second research question, "Which training component is the most effective method of training on the accuracy of ISP-Function data collected?" the research team use the analyses to compare the two separate training protocols to determine differences in effect. Based on the visual analysis, Tau-*U* effect sizes, and results of the multilevel model, there is no evidence that one training component is more effective than the other. While two

participants showed improvements following the Performance Feedback training for both Overall Agreement and Consequence Agreement, and one participant demonstrated improvement following the Behavior Basics training for these variables, all effect sizes were small. Alternatively, three participants demonstrated improvements with small effect sizes following the Behavior Basics training for Disruptive Behavior Agreement, and none showed improvement following Performance Feedback. The multilevel model did not show any statistically significant treatment effects in any phase for any of the three variables indicating there was no significant difference in the effect of the two trainings.

Response to Research Question 3

Finally, in response to the final research question, "Is there a difference between the order in which the training components are implemented on the accuracy of ISP-Function data collected?", the research team evaluated the analyses to determine if there was a difference in effect based on the condition or order in which the treatments were delivered. Based on the visual analysis, Tau-U effect sizes, and results of the multilevel model, there is no evidence that the order in which the trainings are implemented affects the accuracy of ISP-Function data. Participants 4 and 6, both in condition 2, who received the Performance Feedback training followed by the Behavior Basics training, saw larger gains in accuracy across variables. Specifically, based on the visual analysis and Tau-*U* statistic, Participant 4 saw improvements in one phase for Disruptive Behavior Agreement and saw improvements in both phases for Consequence Agreement and Overall Agreement. Participant 6 showed improvements in one phase across all variables. Alternatively, for the participants in condition 1 (Participants 1, 3, and 5), only Participant 3 saw an improvement in one phase in Disruptive Behavior Agreement. However, none of the main effects in the multilevel model were statistically significant based on condition, and all Tau-*U* effect sizes were considered small. Considering these statistically non-

significant findings, and a lack of power given the small sample size, there is no evidence of a significant difference between conditions.

Contributions to the Literature

This study aimed to contribute to the growing body of literature on the training necessary to use of Direct Behavior Ratings and the effectiveness of teacher training components on improving data accuracy. Direct behavior ratings began as a method to collect brief, objective, observational data for screening, and progress monitoring. The defining features of DBR are the preservation of the objectivity of direct observation while calling upon the brevity and simplicity of tools like brief rating scales. DBR has shown promise for assessing disruptive behavior in various populations from kindergarten to eighth grade (e.g., Chafouleas, Kilgus, Jaffery, Riley-Tillman, Welsh & Christ, 2013). Several studies have evaluated the efficacy of various training components and methods on direct behavior rating single item scale (DBR-SIS) accuracy. Chafouleas et al. (2012) reported that completing a comprehensive training package did improve the accuracy of rating academic engagement and disruption but that a comprehensive package may not be more beneficial than a standard training package. However, findings from a study conducted by LeBel and colleagues (2010) suggested that neither the indirect or direct teacher training, which included performance feedback, improved the accuracy of disruptive behavior ratings. By conducting analyses for Disruptive Behavior Agreement in addition to Overall Agreement and Consequence Agreement, this study built upon previous research for DBR-SIS for disruptive behavior. The results of this study indicated little improvements following either training condition for disruptive behavior only. Given the high agreement levels in baseline across participants, this study supports the continued evaluation of a DBR-SIS for disruptive behavior in use in preschool populations with minimal training. Due to the small sample size,

conclusions cannot be drawn regarding the generalizability of these findings. However, this study establishes promise for further research of DBR-SIS in preschool populations.

Less research has been done on using DBR for assessing consequences to determine the function of a behavior. In 2017, Kilgus et al. examined necessary training for undergraduate students to generate accurate consequence ratings. This study's results suggested that training with performance feedback was required to reach accurate ratings of disruptive behavior and consequences using the ISP-Function (Kilgus et al., 2017). Participants who received performance feedback outperformed those in the control group and in training only conditions. As such, in moving forward in research evaluating the ISP-Function in schools, researchers have used a comprehensive training that included performance feedback (e.g., Kilgus et al., 2019). The pilot study used undergraduate students, which is congruent with the teacher population given typical minimum requirements of an undergraduate degree to teach at the elementary level. However, the variation in early childhood educator backgrounds constituted it necessary to reexamine appropriate training to implement this tool in early childhood. Indeed, a 2016 study, which evaluated preschool teacher base knowledge of behavior principles, suggested they may benefit from additional training of functional behavior assessments (Martin, 2016). This study's purpose was to evaluate the degree of teacher training necessary to reach accuracy for both disruptive behavior and consequences. The limitations experienced by the research team prohibited the research team from making conclusive statements regarding the degree of training necessary to use DBR for consequences in early childhood. Implications of this study's limitations are described below, as well as how these limitations can be addressed in future research to ameliorate some of the confounding methodological factors.

Implications for Practice

The difficulties experienced by the research team in obtaining valid data may suggest that the function of preschool disruptive behavior may be best measured with alternative methods. Evidence has shown the effectiveness of this tool in high-frequency behavior (e.g., calling out) and indicates that up to three observations are necessary to collect meaningful and reliable data for ratings of disruptive behavior, adult attention, and peer attention (Kilgus et al., 2019). Alternatively, for escape and access to tangibles and activities, 8-18 observations would be necessary to achieve adequate reliability (Kilgus et al., 2019). The difficulty in predicting students' disruptive behaviors, as experienced by our research team, may present teachers with challenges in collecting the number of data points necessary for reliability. Many preschoolers display episodic behavior problems (e.g., tantrums) that often occur in response to frustration, fatigue, or imposed limit-setting due to poor emotional regulation (Wakschlag et al., 2005). Behavior that is so dependent upon internal emotive responses can be difficult to predict (a necessary factor for planning interval observations). Existing tools such as the Behavior Incident Report (BIR) used within the Pyramid Model may be a more appropriate way to measure disruptive behavior's consequences for students with high intensity and low-frequency behaviors. The BIR is a report that in completed following an incident. Because this is not a direct observation tool, it does not require predetermined observation periods, relying heavily on the predictability of disruptive behavior. Alternatively, this assessment method calls for teachers to reflect directly following an event (e.g., tantrum) on the antecedents, consequences, and setting of a student's disruptive behavior. Collecting series' of BIR's can assist school teams in detecting patterns in setting events or consequences that evoke or maintain problem behavior.

Assessing the rate of disruptive behavior or consequences through DBR may be useful with a smaller portion of preschool students with extremely high-frequency disruptive behavior

in early childhood. While this type of behavior is more uncommon in preschool-aged children than in elementary-aged children, it may still be beneficial for some students. Participant 6's student, for example, had much higher frequency behavior, and throughout the data collection process, 85% of the observations were considered valid. For students like this, DBR may be more appropriate than for students like Participant 2's student who primarily has occasional tantrums and only displayed disruptive behavior in 30% of observations. The implications of the amount of training needed to collect data on the consequences of student behavior accurately should be further researched, given the lack of effects found in this research.

Despite the lack of treatment effects observed across participants, the low levels of Baseline disagreement for the Disruptive Behavior variable suggest that teachers need minimal training to accurately rate disruptive behavior. Based on the multilevel model, the Disruptive Behavior model's intercept was not significant, indicating there were no significant levels of disagreement at Baseline for any of the participants. Indeed, all Baseline disagreement scores for five participants (excluding Participant 6) did not have any scores above 2.3. Four of the six participants (Participants 1, 2, 3, and 4) did not have discrepancies greater than 1.6 points from the true score out of a total of 6. This finding supports the use of Direct Behavior Ratings at the early childhood level to rate the percentage of disruptive behavior. High teacher accuracy scores in Baseline and little change in accuracy following trainings support the implementation of the rating of disruptive behavior via DBR without training. Support for the accuracy in DBR for rating behavior has critical implications for data collection in schools. Disruptive behavior ratings were also largely unaffected by the trainings and levels varied across participants. Because four of the six participants showed highly accurate ratings (with 1-2 points of the true score), many teachers may not need any training to collect this type of data. However, there are apparent individual differences seen in accuracy and between participants indicating some

teachers may need additional training to collect this type of data, or that individual biases may influence their ratings. More research is required to determine factors related to individual differences.

School psychologists can use preschool teachers as data collectors through DBR to appropriately and effectively engage early childhood educators in the consultative process. Teachers collecting classroom data for the frequency of disruptive behavior within as data collectors can be critical tools within the consultative process. Having preschool teachers collect classroom data may be more efficient than if a third-party observer was required to observe the behavior. Using teachers as data collectors can be utilized in each area of the problem-solving process. Initially, these data are useful within the problem identification and problem analysis phases. In problem identification, teams can use teacher-conducted DBR to determine baseline levels of disruptive behavior. Within problem analysis, teams can examine patterns in frequency across days, times and determine potential antecedents (e.g., times of the day behavior occurs at a higher rate). While third-party observers play an essential role, teachers can collect day to day data that may give a more complete picture for decision making and intervention planning.

Engaging teachers in reviewing problem identification and analysis data that they collected may help increase buy-in for the problem-solving process and interventions chosen based on their data. Additionally, teachers can serve as data collectors with DBR during intervention implementation phases to evaluate intervention effectiveness. Ongoing progress monitoring is a critical step in the problem-solving process to ensure that interventions are effective or support changes made to the interventions. School psychologists may best serve to compile the data collected by the teacher and graph it for data interpretation. Engaging teachers in the implementation and evaluation phases, can help build the consultative relationship, and

improve the teacher's understanding in decision-making processes, and increase teacher self-efficacy in working with problem behaviors in the future.

Limitations

Several limitations exist inherently when conducting a single-case design. The first is the lack of generalizability or external validity, given the small sample size. The research team attempted to oversample to account for attrition. Five of the six participants in the study completed the study, with only Participant 2 dropping out following Phase 1. While the research team attempted to recruit a representative sample of early childhood teachers, the results may not be generalizable to preschool teachers in other geographical regions, living in different levels of urbanization, working in other types of preschool settings (e.g., Head Start) or teachers that have different training backgrounds. Further, the small sample size presents difficulty in detecting individual differences in treatment effects that may emerge in large group studies.

The primary limitation of this study was the presence of a significant amount of "incomplete data." As described in the Results section, incomplete data occurred when no disruptive behavior occurred within the observation period. These observations are incomplete because perfect agreement is highly likely, which distinctly skews the agreement data. While the research team anticipated a certain degree of incomplete data, the significant amount of disruptive behavior was unexpected. Across participants, a total of 43% of observations were considered incomplete, with 26% of Baseline observations, 57% of Phase 1 observations, and 56% of Phase 2 observations being incomplete. Participants ranged in the amount of incomplete data with Participant 2 having the most amount of incomplete observations (70%) and Participant 6, having the least amount of incomplete data (16%). The remainder of the participants had between 41-58% of incomplete data points across phases.

Given the importance of observing disruptive behavior during the interval, to avoid the occurrence of incomplete data, the research team implemented a screening procedure to maximize the possibility of observing disruptive behavior. A modified version (see Appendix C) of the Functional Assessment Checklist for Teachers and Staff (FACTS) was utilized to determine appropriate student participants with the teacher. To pass the screening, the teacher had to indicate that the problem behavior occurred every day, at least five or more times per day. Further, within the identifying routines and likelihood of problem behavior, at least one activity during the day had to be rated a 5 or 6, meaning it has a high probability of occurring during that time. This was intended to ensure the behavior was predictable enough to schedule an observation during a high likelihood time and avoid situations in which disruptive behavior was episodic or unpredictable. Of the fourteen students screened, eight students passed the screening. Two of those did not receive parental consent, and the remaining six were the study participants. One limitation of this screening process is the subjective nature of the screening method. Teachers may have overestimated the problem behavior as three of them reported the frequency to be "all the time." Reducing subjectivity in the screening protocol may be a helpful tool in the future and is discussed in the future directions section. Another limitation of this screening procedure was the timing in which screening occurred. The FACTS were completed with the teacher in April, and data were collected during the summer months. The change in the classroom setting's characteristics from the school year to summer presented a methodological challenge for the research team. While the selected students had the same classroom, teachers, and most of the same students as during the year, the number of students in some of the classrooms decreased significantly. In one instance (Participant 5), the student's behavior was rated on the FACTS based on the school year when there are 18-22 students each day with a teacher and an associate. During the summer, when data collection occurred, class size dropped

to an average of 5-6 students with a teacher and an associate. Increasing teacher-student ratio helps by allowing more attention per student and therefore reducing attention-seeking behavior. Further, it will enable the teacher to use greater antecedent control with each of their students (e.g., proximity control) simply by reducing the number of students in the classroom. This significant change in student-teacher ratio occurred in two of the classrooms throughout all observations. The remaining classrooms had considerable variability in teacher-student ratios due to the higher rate of student absences during the summer (e.g., family vacations). Further, the teachers reported that the summers are somewhat less rigorous in terms of academic content. Therefore, another factor hypothesized by several teachers was that there were fewer behavior problems due to the decreased task demands being placed on the child. Decreases in task demand create fewer opportunities for noncompliance and fewer transitions from preferred activities to non-preferred activities, which many of the teachers noted in the screening interview as antecedents to the child's behavior problems.

Future Directions

Given the methodological limitations of this study, several considerations should be made in research designs in furthering this work. In evaluating teacher trainings, specifically for this tool and for direct behavior ratings that require time sampling, it may be more effective to use videos rather than relying on student behavior. As aforementioned, disruptive behavior occurrences within the observation time frame are critical in determining a true level of accuracy. Few disagreements are observed in the absence of disruptive behavior, which was confirmed by the statistical significance of disruptive behavior as a covariate within the multilevel model. Using videos or vignettes may limit the external validity or generalizability to real classroom observation. However, this method would control incomplete observations and allow the research team to examine the effects of teacher training more accurately. Research

using videos and vignettes as a basis for rating has been conducted with success in past studies (e.g., Kilgus et al., 2017). Using videos would allow for a stronger true score to be established by multiple experts or a panel of expert raters. Alternatively, if using real student behavior as the basis for ratings, a more thorough screening method should be employed that includes less subjective methods. This study used a modified version of the Functional Assessment Checklist for Teachers and Staff (FACTS) that relies on teacher reports of the rate and severity of student behavior. Due to restraints in consent procedures, researchers were not allowed to observe students during the screening process to confirm teacher reports of disruptive behavior. Such direct observation by the research team could ensure the behavior is a high enough frequency and occurs predictably enough for reliable observation data. Further, future research teams should take steps to ensure few changes in the environment between screening and data collection.

Considering the use of videos in the research design may also allow for a larger sample size due to decreased challenges, such as finding eligible student participants and obtaining parental consent. The use of videos would also allow researchers to cut down on the amount of time required by participants as this study required daily participation across 4-6 weeks. Large group designs will allow for more power within the study and more reliable conclusions to be drawn from statistical analyses and evaluation of more individual differences. Given the small sample size of this single case design, few conclusions could be drawn about individual differences. Individual differences that may be of particular interest are level and type of education and training considering the varying qualifications of early childhood educators across the county. The teachers in this sample ranged from having an Associate's degree to a Master's degree. Other individual differences that researchers may want to consider is the type of early childhood setting the tool is being implemented in (e.g., private preschool, VPK, HeadStart),

demographics of the child, or types of disruptive behavior being observed. Finally, given this tool showed some promise of being used in the absence of rater training for disruptive behavior, research can be expanded to different populations such as early childhood settings for students with disabilities, public education settings, or HeadStart using a larger group design. However, a primary challenge with using videos is the threat to external validity. Using videos may affect the generalizability of the study to real students in the classroom, particularly in videos that use actors or staged disruptive behavior. Classroom disruptive behavior can be more nuanced and chaotic than may be able to be captured with the videos. Finally, the usage of videos changes the conditions in which teachers are rating students' behavior. Typically, teachers will be engaging in traditional instruction or supervision of their classroom in addition to rating a student. This additional demand cannot be replicated with a video and instead focuses on the teachers' rating skill in a more sterile and controlled environment that does not mimic the full complexities of the demand of the teacher in true implementation. Careful consideration of the benefits of larger sample sizes and pitfalls of generalizability should be made when contemplating using videos in place of student participants.

Future research can expand to evaluate its utility in practice within an early childhood consultative model given the low levels of disagreement for rating disruptive behavior in Baseline. It is unclear how incorporating the teacher into the problem-solving process through the role of a data collector may influence the consultative relationship. Teachers may feel it is an additional responsibility, which could hurt their buy-in in the process. Alternatively, teachers may feel empowered to be active participants and feel an increased involvement within the consultative relationship. Researchers should carefully consider the acceptability of the tool and teacher perceptions of themselves as raters.

Summary

In conclusion, this study has supplemented current research on DBR by investigating the prospect of usage within the Early Childhood setting and the degree of training necessary for implementation. Specifically, this study used a multiple baseline single-case design to examine changes in the accuracy of DBR ratings following various training protocols in 6 early childhood teacher-student dyads. The baseline training protocols consisted of basic exposure to the tool to serve as current practice, as many tools are used in the absence of any rater training. The Behavior Basics training protocol included a didactic training on the basics of behavior, function of behavior, and the importance of using behavior function in intervention development. The Performance Feedback training consisted of a practice component using pre-recorded videos, and participants were provided performance feedback for their responses. The latter two protocols were counterbalanced in participants, so half got the Behavior Basics training first (i.e., condition 1), and half got the Performance Feedback training first (i.e., condition 2).

Teachers interested in the study went through a screening process using the FACTS to determine if a student in their class had appropriate disruptive behavior for the study. To enroll in the study, the student needed to have high-frequency behavior that occurred predictably enough for researchers to rate the student's disruptive behavior. Despite these screening procedures, the research team encountered difficulties in observing necessary amounts of disruptive behavior to determine accuracy in teacher ratings. The research team hypothesized that the percentage of disruptive behavior affected the accuracy ratings, and it was therefore built into the multilevel model as a covariate. The multilevel model indicated that the covariate of the percentage of disruptive behavior was statistically significant across all three variables. These results indicate the need for additional research to determine treatment effects with implications for research methodology.

Analyses were conducted for three separate variables: (1) Overall Agreement, (2) Disruptive Behavior Agreement, and (3) Consequence Agreement. To determine intervention effects, visual analyses were conducted, and the Tau-*U* statistic was calculated to produce an effect size. Finally, a multilevel model was conducted for each of the three variables. The visual analysis and Tau-*U* effect sizes revealed no consistent intervention effects across participants. While there were some increases in accuracy between phases in some participants, all effect sizes were small, and no patterns were detected across participants. The multilevel model showed no statistically significant main effects across variables. This suggests there was no functional relationship between the training protocols and accuracy in ratings using the DBR tool. Further, the lack of main effects indicates there was not a difference in effect between training protocols or based on the order in which the trainings were delivered. However, for the Disruptive Behavior Agreement variable, the intercept was not statistically significant, indicating high levels of agreement within the baseline phase across participants. Indeed, in the visual analysis, few changes were detected within this variable due to floor effects. This has important implications for practice in early childhood. It suggests that teachers can collect accurate data on the percentage of disruptive behavior within an observation in the absence of rater training. Using teachers as data collectors with DBR in early childhood reduces the need for third-party observers for tier 2 behavior problems, can build the consultative relationship, and build buy-in for the teachers in the problem-solving process.

Early detection of developmental delays and psychopathology in early childhood education is an integral piece in addressing the national crisis outlined by the President's New Freedom Commission on Mental Health. So stated, the commission addresses the growing pandemic of mental health concerns that effect children and adolescents alike, leading to negative school outcomes and long-term social outcomes. Widely supported is the emphasis

placed on early identification and intervention. This study aimed to build and expand upon the growing research of DBR and merge with the research being done in early childhood to identify better ways to identify, measure, and treat emotional and behavioral problems.

Collectively, this study suggests early childhood educators may be able to use Direct Behavior Ratings (DBR) for disruptive behavior ratings only in the absence of extensive rater training as agreement levels within baseline phases and across intervention phases generally fell within the acceptable accuracy range (within two points of the true score; Kilgus et al., 2019). DBR may be most useful for students who display an extremely high frequency of disruptive behavior for raters to capture enough data to make meaningful conclusions. However, further research is needed to determine the full utility of DBR in rating the function or consequences of behavior in the early childhood setting to support early childhood educators in addressing behavioral concerns before they intensify and cause lasting academic impact and chronic social-emotional challenges.

References

Ackerman, D. J., & Barnett, W. S. (2005). Prepared for Kindergarten: What Does "Readiness" Mean?, *NIEER Policy Report*.

Anderson, C. M., Rodriguez, B. J., & Campbell, A. (2015). Functional Behavior Assessment in Schools: Current Status and Future Directions. *Journal of Behavioral Education*, *24*(3), 338–371. https://doi.org/10.1007/s10864-015-9226-z

Barbarin, O. A. (2007). Mental health screening of preschool children: Validity and reliability of ABLE. *American Journal of Orthopsychiatry*, *77*(3), 402–418. https://doi.org/10.1037/0002-9432.77.3.402

Aaron C. Barnes, & Jason E. Harlacher. (2008). Clearing the Confusion: Response-to-intervention as a Set of Principles. *Education and Treatment of Children,* 31(3), 417.

Borlow, D., Nock, M., & Hersen, M. (2009). Single case experimental designs. Strategies for Studying Behavior Change 3rd Edition. *Pearson Publishing*.

Barnett, W. S., & Yarosz, D. J. (2007). Who Goes to Preschool and Why Does it Matter?. Preschool Policy Matters, 7.

Barrett, S. B., Bradshaw, C. P., & Lewis-Palmer, T. (2008). Maryland Statewide PBIS Initiative: Systems, Evaluation, and Next Steps. *Journal of Positive Behavior Interventions*, *10*(2), 105–114. https://doi.org/10.1177/1098300707312541

Blair, C., & Raver, C. C. (2015). School Readiness and Self-Regulation: A Developmental Psychobiological Approach. *Annual Review of Psychology*, *66*(1), 711–731. https://doi.org/10.1146/annurev-psych-010814-015221

Blair, K.S. C., Fox, L., & Lentini, R. (2010). Use of Positive Behavior Support to Address the Challenging Behavior of Young Children Within a Community Early Childhood Program. *Topics in Early Childhood Special Education*, *30*(2), 68–79. https://doi.org/10.1177/0271121410372676

Borgmeier, C., Loman, S. L., & Strickland-Cohen, M. K. (2017). ABC Tracker: Increasing Teacher Capacity for Assessing Student Behavior. *Beyond Behavior*, *26*(3), 113–123. https://doi.org/10.1177/1074295617728513

Bowman, B. T., Donovan, S., & Burns, M. S. (2000). Eager to learn: Educating our preschoolers: Executive summary. National Academy Press.

Bradley, R. H. (2003). National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network. Child care and common communicable illnesses in children aged 37 to 54 months. *Arch Pediatr Adolesc Med*, *157*(2), 196-200.

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of School-Wide Positive Behavioral Interventions and Supports on Child Behavior Problems. *Pediatrics*, *130*(5), e1136–e1145. https://doi.org/10.1542/peds.2012-0243

Bradshaw, C. P., Koth, C. W., Thornton, L. A., & Leaf, P. J. (2009). Altering School Climate through School-Wide Positive Behavioral Interventions and Supports: Findings from a Group-Randomized Effectiveness Trial. *Prevention Science*, *10*(2), 100–115. https://doi.org/10.1007/s11121-008-0114-9

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the Effects of Schoolwide Positive Behavioral Interventions and Supports on Student Outcomes: Results From a Randomized Controlled Effectiveness Trial in Elementary Schools. *Journal of Positive Behavior Interventions*, *12*(3), 133–148. https://doi.org/10.1177/1098300709334798

Briggs-Gowan, M. J., Carter, A. S., Skuban, E. M., & Horwitz, S. M. (2001). Prevalence of Social-Emotional and Behavioral Problems in a Community Sample of 1- and 2-Year-Old Children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(7), 811–819. https://doi.org/10.1097/00004583-200107000-00016

Bruhn, A., Barron, S., Fernando, J., & Balint-Langel, K. (2018). Extending the direct behavior rating: An examination of schoolwide behavior ratings and academic engagement. *Journal of Positive Behavior Interventions*, 20(1), 31-42.

Bruni, T. P., Drevon, D., Hixson, M., Wyse, R., Corcoran, S., & Fursa, S. (2017). The Effect of Functional Behavior Assessment on School-Based Interventions: A Meta-Analysis of Single-Case Research: Functional Behavior Assessment in Schools. *Psychology in the Schools*, *54*(4), 351–369. https://doi.org/10.1002/pits.22007

Bulotsky-Shearer, R. J., Dominguez, X., & Bell, E. R. (2012). Preschool classroom behavioral context and school readiness outcomes for low-income children: A multilevel examination of child- and classroom-level influences. *Journal of Educational Psychology*, *104*(2), 421–438. https://doi.org/10.1037/a0026301

Caldarella, P., Shatzer, R. H., Gray, K. M., Young, K. R., & Young, E. L. (2011). The Effects of School-wide Positive Behavior Support on Middle School Climate and Student Outcomes. *RMLE Online*, *35*(4), 1–14. https://doi.org/10.1080/19404476.2011.11462087

Campbell, A., & Anderson, C. M. (2008). Enhancing Effects of Check-in/Check-out with Function-Based Support. *Behavioral Disorders*, *33*(4), 233–245. https://doi.org/10.1177/019874290803300404

Chafouleas, S. M. (2011). Direct Behavior Rating: A Review of the Issues and Research in Its Development. *Education and Treatment of Children*, *34*(4), 575–591. https://doi.org/10.1353/etc.2011.0034

Chafouleas, S. M., Christ, T. J., Riley-Tillman, T. C., Briesch, A., & Chanese, J. A. M. (2007). Generalizability and Dependability of Direct Behavior Ratings to Assess Social Behavior of Preschoolers, *School Psychology Review*, 36(1), 63-79.

Chafouleas, S. M., Kilgus, S. P., Jaffery, R., Riley-Tillman, T. C., Welsh, M., & Christ, T. J. (2013). Direct behavior rating as a school-based behavior screener for elementary and middle grades. *Journal of School Psychology*, 51(3), 367-385.

Chafouleas, S. M., Kilgus, S. P., Riley-Tillman, T. C., Jaffery, R., & Harrison, S. (2012). Preliminary evaluation of various training components on accuracy of Direct Behavior Ratings. *Journal of School Psychology*, *50*(3), 317–334. https://doi.org/10.1016/j.jsp.2011.11.007

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. *The Journal of Special Education*, *42*(2), 108–126. https://doi.org/10.1177/0022466907313349

Christ, T. J., Riley-Tillman, T. C., & Chafouleas, S. M. (2009). Foundation for the Development and Use of Direct Behavior Rating (DBR) to Assess and Evaluate Student Behavior. *Assessment for Effective Intervention*, *34*(4), 201–213. https://doi.org/10.1177/1534508409340390

Christner, R. W., Forrest, E., Morley, J., & Weinstein, E. (2007). Taking Cognitive-Behavior Therapy to School: A School-Based Mental Health Approach. *Journal of Contemporary Psychotherapy*, *37*(3), 175–183. https://doi.org/10.1007/s10879-007-9052-2

Committee for Children. (2011). Second Step: Social-emotional skills for early learning.

Cook, C. R., Frye, M., Slemrod, T., Lyon, A. R., Renshaw, T. L., & Zhang, Y. (2015). An integrated approach to universal prevention: Independent and combined effects of PBIS

and SEL on youths' mental health. *School Psychology Quarterly*, *30*(2), 166–183. https://doi.org/10.1037/spq0000102

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). Applied behavior analysis. *Pearson Publishing.*

Crone, D. A., Hawken, L. S., & Horner, R. H. (2010). *Responding to problem behavior in schools: The behavior education program*. Guilford Press.

DiPerna, J. C. (2006). Academic enablers and student achievement: Implications for assessment and intervention services in the schools. *Psychology in the Schools*, *43*(1), 7–17. https://doi.org/10.1002/pits.20125

Doll, B., Cummings, J. A., & Chapla, B. A. (2008). Best practices in population-based school mental health services. *Best practices in school psychology V*, *4*, 1333-1347.

Doll, B., & Cummings, J. (2008). Why population-based services are essential for school mental health, and how to make them happen in your school. *Transforming school mental health services: Population-based approaches to promoting the competency and wellness of children*, 1-20. Corwin Press.

Drummond, T. (1994). The student risk screening scale (SRSS). *Grants Pass, OR: Josephine County Mental Health Program*.

Dufrene, B. A., Doggett, R. A., Henington, C., & Watson, T. S. (2007). Functional Assessment and Intervention for Disruptive Classroom Behaviors in Preschool and Head Start Classrooms. *Journal of Behavioral Education*, *16*(4), 368–388. https://doi.org/10.1007/s10864-007-9048-8

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Dunlap, G., & Fox, L. (2011). Function-Based Interventions for Children With Challenging Behavior. *Journal of Early Intervention*, *33*(4), 333–343. https://doi.org/10.1177/1053815111429971

Dunlap, G., Kern, L., Deperczel, M., Clarke, S., Wilson, D., Childs, K.E., White, R. and Falk, G.D (2018). Republication of "Functional Analysis of Classroom Variables for Students with Emotional and Behavioral Disorders." *Behavioral Disorders*, *43*(2), 322–336. https://doi.org/10.1177/0198742917753157

Dunlap, G., Strain, P.S., Fox, L., Carta, J.J., Conroy, M., Smith, B.J., Kern, L., Hemmeter, M.L., Timm, M.A., McCart, A. and Sailor, W., (2006). Prevention and intervention with young children's challenging behavior: Perspectives regarding current knowledge. *Behavioral Disorders*, 29–45.

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: Social and Emotional Learning. *Child Development*, *82*(1), 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

Egger, H. L., & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry*, *47*(3–4), 313–337. https://doi.org/10.1111/j.1469-7610.2006.01618.x

Eyberg, S. M., & Pincus, D. (1999). *ECBI & SESBI-R: Eyberg child behavior inventory and Sutter-Eyberg student behavior inventory-revised: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Ferron, J., & Jones, P. K. (2006). Tests for the Visual Analysis of Response-Guided Multiple-Baseline Data. *The Journal of Experimental Education*, *75*(1), 66–81. https://doi.org/10.3200/JEXE.75.1.66-81

Ferron, J. M., Bell, B. A., Hess, M. R., Rendina-Gobioff, G., & Hibbard, S. T. (2009). Making treatment effect inferences from multiple-baseline data: The utility of multilevel modeling approaches. *Behavior Research Methods*, *41*(2), 372–384. https://doi.org/10.3758/BRM.41.2.372

Fox, L., Dunlap, G., Hemmeter, M. L., Joseph, G. E., & Strain, P. S. (2003). The Teaching Pyramid: A Model for Supporting Social Competence and Preventing Challenging Behavior in Young Children. *Young Children*, *58*(4), 48-52.

Fox, L., & Little, N. (2001). Starting Early: Developing School- Wide Behavior Support in a Community Preschool. *Journal of Positive Behavior Interventions*, *3*(4), 251–254. https://doi.org/10.1177/109830070100300406

Freeman, J., Simonsen, B., McCoach, D. B., Sugai, G., Lombardi, A., & Horner, R. (2016). Relationship Between School-Wide Positive Behavior Interventions and Supports and Academic, Attendance, and Behavior Outcomes in High Schools. *Journal of Positive Behavior Interventions*, *18*(1), 41–51. https://doi.org/10.1177/1098300715580992

Furniss, T., Beyer, T., & Guggenmos, J. (2006). Prevalence of behavioural and emotional problems among six-years-old preschool children: Baseline results of a prospective longitudinal study. *Social Psychiatry and Psychiatric Epidemiology*, *41*(5), 394–399. https://doi.org/10.1007/s00127-006-0045-3

Gadow, K. D., & Sprafkin, J. (1997). *Child Symptom Inventory 4: CSI*. Checkmate Plus.

Gage, N. A., Lewis, T. J., & Stichter, J. P. (2012). Functional Behavioral Assessment-Based Interventions for Students with or at Risk for Emotional and/or Behavioral Disorders in School: A Hierarchical Linear Modeling Meta-Analysis. *Behavioral Disorders*, *37*(2), 55–77. https://doi.org/10.1177/019874291203700201

Garmezy, N., Masten, A. S., & Tellegen, A. (1984). The Study of Stress and Competence in Children: A Building Block for Developmental Psychopathology, *Child Development*, 97-111.

Garver, K. (2020). NIEER Pre-K Data Snapshot: State Pre-K Policies--Salary Parity Varies with Teacher Qualifications and Setting. *National Institute for Early Education Research*.

Goodman, R. (1997) Strengths and Difficulties Questionnaire (SDQ).

Gresham, F. M. (2011). Social Behavioral Assessment and Intervention: Observations and Impressions. *School Psychology Review*, *40*(2), 10.

Gresham, F. M., Cook, C. R., Crews, S. D., & Kern, L. (2004). Social Skills Training for Children and Youth with Emotional and Behavioral Disorders: Validity Considerations and Future Directions. *Behavioral Disorders*, *30*(1), 32–46. https://doi.org/10.1177/019874290403000101

Gresham, F. M., Elliott, S. N., Vance, M. J., & Cook, C. R. (2011). Comparability of the Social Skills Rating System to the Social Skills Improvement System: Content and psychometric comparisons across elementary and secondary age levels. *School Psychology Quarterly*, *26*(1), 27–44. https://doi.org/10.1037/a0022662

Gresham, F. M., Sugai, G., & Horner, R. H. (2001). Interpreting Outcomes of Social Skills Training for Students with High-Incidence Disabilities. *Exceptional Children*, *67*(3), 331–344. https://doi.org/10.1177/001440290106700303

Hawken, L. S. (2006). School psychologists as leaders in the implementation of a targeted intervention: The Behavior Education Program. *School Psychology Quarterly*, *21*(1), 91.

Hawken, L. S., Adolphson, S. L., MacLeod, K. S., Schumann, J., Sailor, W., Sugai, G., Horner, R.H. & Dunlap, G. (2009). Handbook of positive behavior support. Springer.

Hawken, L. S., & Horner, R. H. (2003). Evaluation of a Targeted Intervention Within a Schoolwide System of Behavior Support. *Journal of Behavioral Education*, 12(3), 225-240.

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*(4), 551–574. https://doi.org/10.1353/etc.2011.0031

Hesney, J. (2012). Teacher knowledge and perceptions of the school-based functional behavioral assessment (Doctoral dissertation, Fairleigh Dickinson University).

Hindman, A. H., Skibbe, L. E., Miller, A., & Zimmerman, M. (2010). Ecological contexts and early learning: Contributions of child, family, and classroom factors during Head Start, to literacy and mathematics growth through first grade. *Early Childhood Research Quarterly*, *25*(2), 235–250. https://doi.org/10.1016/j.ecresq.2009.11.003

Horner, R.H., Kincaid, D., Sugai, G., Lewis, T., Eber, L., Barrett, S., Dickey, C.R., Richter, M., Sullivan, E., Boezio, C. and Algozzine, B., (2014). Scaling Up School-Wide Positive Behavioral Interventions and Supports: Experiences of Seven States With Documented Success. *Journal of Positive Behavior Interventions*, *16*(4), 197–208. https://doi.org/10.1177/1098300713503685

Huaqing Qi, C., & Kaiser, A. P. (2003). Behavior Problems of Preschool Children From Low-Income Families: Review of the Literature. *Topics in Early Childhood Special Education*, *23*(4), 188–216. https://doi.org/10.1177/02711214030230040201

Iwata, B. A., DeLeon, I. G., & Roscoe, E. M. (2013). Reliability and Validity of the Functional Analysis Screening Tool: Functional Analysis Screening Tool. *Journal of Applied Behavior Analysis*, *46*(1), 271–284. https://doi.org/10.1002/jaba.31

Johnson, A. H., Miller, F. G., Chafouleas, S. M., Welsh, M. E., Chris Riley-Tillman, T., & Fabiano, G. (2016). Evaluating the technical adequacy of DBR-SIS in tri-annual behavioral screening: A multisite investigation. *Journal of School Psychology*, *54*, 39–57. https://doi.org/10.1016/j.jsp.2015.10.001

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

Justice, L. M., Mashburn, A. J., Hamre, B. K., & Pianta, R. C. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly*, *23*(1), 51–68. https://doi.org/10.1016/j.ecresq.2007.09.004

Kauffman, J. M., & Hallahan, D. P. (Eds.). (2011). *Handbook of special education*. New York: Routledge.

Kaufmann, R., & Wischmann, A. L. (1999). Communities supporting the mental health of young children and their families. *Where children live: Solutions for serving young children and their families*, 175-210.

Kamphaus, R. W., & Reynolds, C. R. (2015). Behavior assessment system for children—third edition (BASC-3): Behavioral and emotional screening system (BESS). *Bloomington, MN: Pearson*.

Kazdin, A. E. (1982). The token economy: A decade later. *Journal of Applied Behavior Analysis*, *15*(3), 431-445.

Kelm, J. L., & McIntosh, K. (2012). Effects of school-wide positive behavior support on teacher self-efficacy. *Psychology in the Schools*, *49*(2), 137–147. https://doi.org/10.1002/pits.20624

Kelm, J. L., McIntosh, K., & Cooley, S. (2014). Effects of Implementing School-Wide Positive Behavioural Interventions and Supports on Problem Behaviour and Academic

Achievement in a Canadian Elementary School. *Canadian Journal of School Psychology*, *29*(3), 195–212. https://doi.org/10.1177/0829573514540266

Kilgus, S. P., Chafouleas, S. M., Riley-Tillman, T. C., & von der Embse, N. P. (2014). Social, academic, and emotional behavior risk screener (SAEBRS). *Minneapolis, MN: Theodore J. Christ & Colleagues*.

Kilgus, S. P., Kazmerski, J. S., Taylor, C. N., & von der Embse, N. P. (2017). Use of direct behavior ratings to collect functional assessment data. *School Psychology Quarterly*, *32*(2), 240–253. https://doi.org/10.1037/spq0000156

Kilgus, S. P., Riley-Tillman, T. C., Stichter, J. P., Schoemann, A. M., & Bellesheim, K. (2016). Reliability of Direct Behavior Ratings – Social Competence (DBR-SC) data: How many ratings are necessary? *School Psychology Quarterly*, *31*(3), 431–442. https://doi.org/10.1037/spq0000128

Kilgus, S. P., von der Embse, N. P., Eklund, K., Izumi, J., Peet, C., Meyer, L., & Taylor, C. N. (2019). Reliability, validity, and accuracy of the intervention selection profile—Function: A brief functional assessment tool. *School Psychology*, *34*(5), 531.

Kircher, K. L. S. (2009). Functional behavioral assessment in schools: Teacher knowledge, perspectives, and discipline referral practices (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses Global. (Order No. 3354430)

Klingbeil, D. A., Dart, E. H., & Schramm, A. L. (2018). A Systematic Review of Function-Modified Check-In/Check-Out. *Journal of Positive Behavior Interventions*, 109830071877803. https://doi.org/10.1177/1098300718778032

Koehler, M. J., & Levin, J. R. (1998). Regulated randomization: A potentially sharper analytical tool for the multiple-baseline design. *Psychological Methods*, *3*(2), 206–217. https://doi.org/10.1037/1082-989X.3.2.206

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. *What works clearinghouse.*

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-Case Intervention Research Design Standards. *Remedial and Special Education*, *34*(1), 26–38. https://doi.org/10.1177/0741932512452794

Lane, K. L., Capizzi, A. M., Fisher, M. H., & Ennis, R. P. (2012). Secondary Prevention Efforts at the Middle School Level: An Application of the Behavior Education Program. *Education and Treatment of Children*, *35*(1), 51–90. https://doi.org/10.1353/etc.2012.0002

Lane, K. L., & Menzies, H. M. (2005). Teacher-Identified Students with and without Academic and Behavioral Concerns: Characteristics and Responsiveness. *Behavioral Disorders*, *31*(1), 65–83. https://doi.org/10.1177/019874290503100103

Lane, K. L., Oakes, W. P., Ennis, R. P., & Hirsch, S. E. (2014). Identifying Students for Secondary and Tertiary Prevention Efforts: How Do We Determine Which Students Have Tier 2 and Tier 3 Needs? *Preventing School Failure: Alternative Education for Children and Youth*, *58*(3), 171–182. https://doi.org/10.1080/1045988X.2014.895573

Lane, K., Parks, R. J., Robertson Kalberg, J., & Carter, E. W. (2007). Systematic Screening at the Middle School Level: Score Reliability and Validity of the Student Risk Screening Scale. *Journal of Emotional and Behavioral Disorders*, *15*(4), 209-222.

Lavigne, J. V., LeBailly, S. A., Hopkins, J., Gouze, K. R., & Binns, H. J. (2009). The Prevalence of ADHD, ODD, Depression, and Anxiety in a Community Sample of 4-Year-Olds. *Journal of Clinical Child & Adolescent Psychology*, *38*(3), 315–328. https://doi.org/10.1080/15374410902851382

LeBel, T. J., Kilgus, S. P., Briesch, A. M., & Chafouleas, S. (2010). The impact of training on the accuracy of teacher-completed direct behavior ratings (DBRs). *Journal of Positive Behavior Interventions*, *12*(1), 55-63.

LeBuffe, P. A., & Naglieri, J. A. (1999). The Devereux early childhood assessment. *Lewisville, NC: Kaplan Press Publishing*.

Lehr, C. A., & McComas, J. (2005). Students with emotional/behavioral disorders: Promoting positive outcomes. *Impact: Feature Issue on Fostering Success in School and beyond for Students with Emotional and Behavioral Disorders*, *18*. Retrieved from http://ncfy.acf.hhs.gov/sites/default/files/docs/13749-Students_with_Emotional_Behavioral_Disorders.pdf

Linver, M. R., Brooks-Gunn, J., & Kohen, D. E. (2002). Family processes as pathways from income to young children's development. *Developmental Psychology*, *38*(5), 719–734. https://doi.org/10.1037//0012-1649.38.5.719

Lloyd, B. P., Weaver, E. S., & Staubitz, J. L. (2017). Classroom-Based Strategies to Incorporate Hypothesis Testing in Functional Behavior Assessments. *Beyond Behavior*, *26*(2), 48–56. https://doi.org/10.1177/1074295617711145

March, R. E., & Horner, R. H. (2002). Feasibility and Contributions of Functional Behavioral Assessment in Schools. *Journal of Emotional and Behavioral Disorders*, *10*(3), 158–170. https://doi.org/10.1177/10634266020100030401

March, R. E., Horner, R. H., Lewis-Palmer, T., Brown, D., Crone, D., Todd, A. W., & Carr, E. (2000). Functional assessment checklist for teachers and staff (FACTS). *Eugene, OR: Educational and Community Supports*.

Martens, B. K., DiGennaro, F. D., Reed, D. D., Szczech, F. M., & Rosenthal, B. D. (2008). Contingency Space Analysis: An Alternative Method for Identifying Contingent

Relations from Observational Data. *Journal of Applied Behavior Analysis*, *41*(1), 69–81. https://doi.org/10.1901/jaba.2008.41-69

Martin, K. B. (2016). Knowledge and Perception of Functional Behavioral Assessment Among Preschool teachers and Early Interventionists, (Doctoral dissertation, Middle Tennessee State University).

Masten, A. S., Desjardins, C. D., McCormick, C. M., Kuo, S. I.-C., & Long, J. D. (2010). The significance of childhood competence and problems for adult success in work: A developmental cascade analysis. *Development and Psychopathology*, *22*(03), 679–694. https://doi.org/10.1017/S0954579410000362

Masten, A.S., Roisman, G.I., Long, J.D., Burt, K.B., Obradović, J., Riley, J.R., Boelcke-Stennes, K. and Tellegen, A., (2005). Developmental Cascades: Linking Academic Achievement and Externalizing and Internalizing Symptoms Over 20 Years. *Developmental Psychology*, *41*(5), 733–746. https://doi.org/10.1037/0012-1649.41.5.733

McCarney, S. (1994). Early Childhood Behavior Scale (ECBS). *Columbia, MO, Hawthorne Educational Services Inc.*

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology*, *43*(4), 947–959. https://doi.org/10.1037/0012-1649.43.4.947

McDougal, J. L., Bardos, A., & Meier, S. T. (2011). *Behavior Intervention Monitoring Assessment System: Technical Manual*. Multi-Health Systems Incorporated.

McIntosh, K., Brown, J. A., & Borgmeier, C. J. (2008). Validity of Functional Behavior Assessment Within a Response to Intervention Framework: Evidence, Recommended

Practice, and Future Directions. *Assessment for Effective Intervention*, *34*(1), 6–14. https://doi.org/10.1177/1534508408314096

McIntosh, K., Campbell, A. L., Carter, D. R., & Rossetto Dickey, C. (2009). Differential Effects of a Tier Two Behavior Intervention Based on Function of Problem Behavior. *Journal of Positive Behavior Interventions*, *11*(2), 82–93. https://doi.org/10.1177/1098300708319127

McLeod, B. D., Sutherland, K. S., Martinez, R. G., Conroy, M. A., Snyder, P. A., & Southam-Gerow, M. A. (2017). Identifying Common Practice Elements to Improve Social, Emotional, and Behavioral Outcomes of Young Children in Early Childhood Classrooms. *Prevention Science*, *18*(2), 204–213. https://doi.org/10.1007/s11121-016-0703-y

Merikangas, K.R., He, J.P., Burstein, M., Swanson, S.A., Avenevoli, S., Cui, L., Benjet, C., Georgiades, K. and Swendsen, J.. (2010). Lifetime Prevalence of Mental Disorders in U.S. Adolescents: Results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *Adolescent Psychiatry*, *49*(10), 10.

Merrell, K. (2003). The Preschool and Kindergarten Behavior Scales-Second Edition (PKBS-2). *Austin, TX: PRO-ED Inc.*

Miller, F. G., Chafouleas, S. M., Riley-Tillman, T. C., & Fabiano, G. A. (2014). Teacher Perceptions of the Usability of School-Based Behavior Assessments. *Behavioral Disorders*, *39*(4), 201–210. https://doi.org/10.1177/019874291303900405

National Research Council, & National Research Council. (2009). Institute of medicine. *From neurons to neighborhoods: The science of early childhood development*, 146-149.

Newton, S. J., Horner, R. H., Algozzine, R. F., Todd, A. W., & Algozzine, K. M. (2009). Using a problem-solving model to enhance data-based decision making in schools. In Handbook of positive behavior support (pp. 551-580). Springer, Boston, MA.

Noltemeyer, A., Petrasek, M., Stine, K., Palmer, K., Meehan, C., & Jordan, E. (2018). Evaluating and Celebrating PBIS Success: Development and Implementation of Ohio's PBIS Recognition System. *Journal of Applied School Psychology*, *34*(3), 215–241. https://doi.org/10.1080/15377903.2017.1381659

Obradović, J., Burt, K. B., & Masten, A. S. (2009). Testing a Dual Cascade Model Linking Competence and Symptoms Over 20 Years from Childhood to Adulthood. *Journal of Clinical Child & Adolescent Psychology*, *39*(1), 90–102. https://doi.org/10.1080/15374410903401120

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining Nonoverlap and Trend for Single-Case Research: Tau-*U*. *Behavior Therapy*, *42*(2), 284–299. https://doi.org/10.1016/j.beth.2010.08.006

Pas, E. T., & Bradshaw, C. P. (2012). Examining the Association Between Implementation and Outcomes: State-wide Scale-up of School-wide Positive Behavior Intervention and Supports. *The Journal of Behavioral Health Services & Research*, *39*(4), 417–433. https://doi.org/10.1007/s11414-012-9290-2

Pianta, R. C., Cox, M. J., & National Center for Early Development & Learning (U.S.) (Eds.). (1999). *The transition to kindergarten*. Baltimore: P.H. Brookes Pub.

Reardon, S. F. (2013). No rich child left behind. *New York Times*, *4*(28), 13.

Regenstein, E., Connors, M., Romero-Jurado, R., & Weiner, J. (2017). Uses and Misuses of Kindergarten Readiness Assessment Results, *Ounce Policy Conversations, 6(11)*. .

Reinke, W. M., Herman, K. C., Petras, H., & Ialongo, N. S. (2008). Empirically Derived Subtypes of Child Academic and Behavior Problems: Co-Occurrence and Distal Outcomes. *Journal of Abnormal Child Psychology*, *36*(5), 759–770. https://doi.org/10.1007/s10802-007-9208-2

Reinke, W. M., Herman, K. C., & Stormont, M. (2013). Classroom-level positive behavior supports in schools implementing SW-PBIS identifying areas for enhancement. *Journal of Positive Behavior Interventions*, *15*(1), 39–50.

Richardson, R. C., Tolson, H., Huang, T.-Y., & Lee, Y.-H. (2009). Character Education: Lessons for Teaching Social and Emotional Competence. *Children & Schools*, *31*(2), 71–78. https://doi.org/10.1093/cs/31.2.71

Riley-Tillman, T. C., Chafouleas, S. M., Sassu, K. A., Chanese, J. A. M., & Glazer, A. D. (2008). Examining the Agreement of Direct Behavior Ratings and Systematic Direct Observation Data for On-Task and Disruptive Behavior. *Journal of Positive Behavior Interventions*, *10*(2), 136–143. https://doi.org/10.1177/1098300707312542

Roeser, R. W., Eccles, J. S., & Sameroff, A. J. (1998). Academic and emotional functioning in early adolescence: Longitudinal relations, patterns, and prediction by experience in middle school. *Development and Psychopathology*, *10*(2), 321–352. https://doi.org/10.1017/S0954579498001631

Simonsen, B., Eber, L., Black, A. C., Sugai, G., Lewandowski, H., Sims, B., & Myers, D. (2012). Illinois Statewide Positive Behavioral Interventions and Supports: Evolution and Impact on Student Outcomes Across Years. *Journal of Positive Behavior Interventions*, *14*(1), 5–16. https://doi.org/10.1177/1098300711412601

Soles, T., Bloom, E. L., Heath, N. L., & Karagiannakis, A. (2008). An exploration of teachers' current perceptions of children with emotional and behavioural difficulties. *Emotional and Behavioural Difficulties*, *13*(4), 275–290. https://doi.org/10.1080/13632750802442201

South, B. (2017). *Paraprofessionals' use of Direct Behavior Rating for collecting Functional Behavior Assessment data: Agreement and Acceptability* (Doctoral dissertation).

Sprague, J., & Walker, H. (2000). Early Identification and Intervention for Youth with Antisocial and Violent Behavior. *Exceptional Children*, *66*(3), 367–379. https://doi.org/10.1177/001440290006600307

Sugai, G., & Horner, R. (2002). The Evolution of Discipline Practices: School-Wide Positive Behavior Supports. *Child & Family Behavior Therapy*, *24*(1–2), 23–50. https://doi.org/10.1300/J019v24n01_03

Suldo, S. M. (2016). *Promoting student happiness: Positive psychology interventions in schools*. Guilford Publications.

Sutherland, K. S., Conroy, M., Abrams, L., & Vo, A. (2010). Improving Interactions between teachers and young children with problem behavior: A strengths-based approach. *Exceptionality*, 18(2), 70 – 81.

Sutherland, K. S., Conroy, M. A., Algina, J., Ladwig, C., Jessee, G., & Gyure, M. (2018). Reducing child problem behaviors and improving teacher-child interactions and relationships: A randomized controlled trial of BEST in CLASS. Early Childhood Research Quarterly, 42, 31-43.

Trussell, R. P., Lewis, T. J., & Raynor, C. (2016). The Impact of Universal Teacher Practices and Function-based Behavior Interventions on the Rates of Problem Behaviors among At-risk Students. *Education and Treatment of Children*, *39*(3), 261–282. https://doi.org/10.1353/etc.2016.0012

US Department of Health and Human Services. (2018). Head Start program facts: Fiscal year 2017. Administration for Children and Families, Office of Head Start.

Vo, A. K., Sutherland, K. S., & Conroy, M. A. (2012). Best in class: A classroom-based model for ameliorating problem behavior in early childhood settings. *Psychology in the Schools*, 49(5), 402-415.

Volpe, R. J., Chaffee, R. K., Yeung, T. S., & Briesch, A. M. (2020). Initial Development of Multi-item Direct Behavior Rating Measures of Academic Enablers. School Mental Health, 12(1), 77-87.

Walker, H., Feil, E.G., Frey, A., Small, J., Seeley, J., Golly, A., Crosby, S., Lee, J., Forness, S., Sprick, M. and Coughlin, C. (2018). First Step Next: An Updated Version of the First Step to Success Early Intervention Program. Grantee Submission, 3(1), 89-109.

Walker, H. M., Nishioka, V. M., Zeller, R., Severson, H. H., & Feil, E. G. (2000). Causal Factors and Potential Solutions for the Persistent Underidentification of Students Having Emotional or Behavioral Disorders in the Context of Schooling. *Assessment for Effective Intervention*, *26*(1), 29–39. https://doi.org/10.1177/073724770002600105

Walker, V. L., Chung, Y.-C., & Bonnet, L. K. (2018). Function-Based Intervention in Inclusive School Settings: A Meta-Analysis. *Journal of Positive Behavior Interventions*, *20*(4), 203–216. https://doi.org/10.1177/1098300717718350

Webster-Stratton, C., Mostyn, D., & Marie, J. S. (2005). *The Incredible Years: A trouble-shooting guide for parents of children aged 2-8 years*. Seattle, WA: Incredible years.

Zarcone, J. R., Rodgers, T. A., Iwata, B. A., Rourke, D. A., & Dorsey, M. F. (1991). Reliability analysis of the motivation assessment scale: A failure to replicate. *Research in Developmental Disabilities*, *12*(4), 349–360. https://doi.org/10.1016/0891-4222(91)90031-M

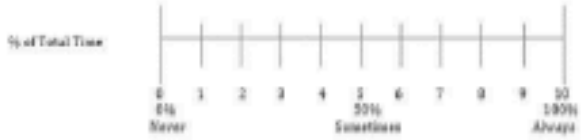
Appendix A: Intervention Selection Profile-Function (ISP-Function)

Date:	Student:	Activity Description:
Time:	Rater:	

Directions: Place a mark along the line that best reflects the percentage of total time the student exhibited the problem behavior.

Disruptive behavior is a student action that interrupts regular school or classroom activity. For example, out of seat, fidgeting, playing with objects, acting aggressively, talking/yelling about things that are unrelated to classroom instruction.

Disruptive Behavior:



Directions: Place a mark along the line that best reflects the percentage of problem behaviors that were followed by each consequence.

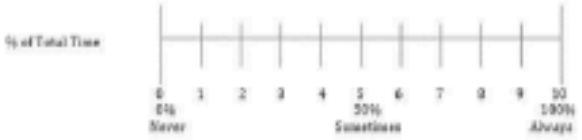
Adult Attention: Positive, negative, or neutral adult reaction that can be either verbal or nonverbal. Examples: reprimand, redirection to work, praise, discussion, high-fives, or shushing.

Peer Attention: Positive, negative, or neutral peer reaction that can be either verbal or nonverbal. Examples: talking, laughing, arguing, high-fives, hitting, kicking, or yelling.

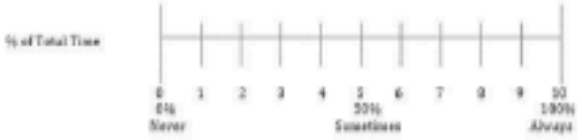
Escape/Avoidance: Removal or avoidance of task, activity, or performance expectations. Examples: removal of academic materials, allowance to delay task completion, permission to leave room, or elimination of task demands.

Access to Tangibles or Activities: Acquisition of items or activities. Examples: toys, food, prizes, games, preferred tasks, sleep, technology, or homework pass.

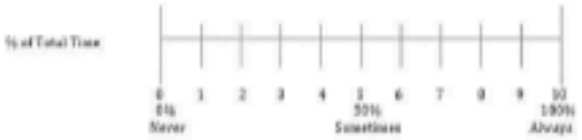
Adult Attention:



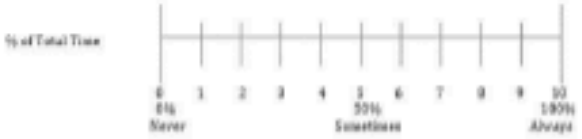
Peer Attention:



Escape/Avoidance:



Access to Tangibles/Activities:



Appendix B: Systematic Direct Observation (SDO) Form

Systematic Direct Observation
15-min, 10-sec intervals

Student: _____ **Start Time:** _____

Date: _____ **End Time:** _____

Instructions: Prior to conducting observations, construct an operational definition of the problem behavior to be observed. Once constructed, insert that definition into the first page of this form. When conducting observations, record whether problem behaviors were observed (+) or not (-) at any point within each 10-sec interval. If the behavior was observed, please also record the consequence that followed the behavior. Specifically, record a '+' in relation to each observed consequence.

Adult Attention (AA): Positive, negative, or neutral adult reaction that can be either verbal or nonverbal.

Peer Attention (PA): Positive, negative, or neutral peer reaction that can be either verbal or nonverbal.

Escape/Avoidance (EA): Removal or avoidance of task, activity, or performance expectations.

Access to Tangibles or Activities (ATA): Acquisition of items or activities.

	0:10	0:20	0:30	0:40	0:50	1:00	1:10	1:20	1:30	1:40	1:50	2:00
PB												
AA												
PA												
EA												
ATA												

	2:10	2:20	2:30	2:40	2:50	3:00	3:10	3:20	3:30	3:40	3:50	4:00
PB												
AA												
PA												
EA												
ATA												

	4:10	4:20	4:30	4:40	4:50	5:00	5:10	5:20	5:30	5:40	5:50	6:00
PB												
AA												
PA												
EA												
ATA												

	6:10	6:20	6:30	6:40	6:50	7:00	7:10	7:20	7:30	7:40	7:50	8:00
PB												
AA												
PA												
EA												
ATA												

	8:10	8:20	8:30	8:40	8:50	9:00	9:10	9:20	9:30	9:40	9:50	10:00
PB												
AA												
PA												
EA												
ATA												

%PB: _____ %AA: _____ %PA: _____ %EA: _____ %ATA: _____

Appendix C: Modified Functional Assessment Checklist for Teachers and Staff (FACTS)

Functional Assessment Checklist for Teachers and Staff (FACTS-Part A)

Student/ Grade: _____ Date: _____
Interviewer: _____ Respondent(s): _____

Problem Behavior(s): Identify problem behaviors

___ Tardy	___ Fight/physical Aggression	___ Disruptive	___ Theft
___ Unresponsive	___ Inappropriate Language	___ Insubordination	___ Vandalism
___ Withdrawn	___ Verbal Harassment	___ Work not done	___ Other _____
	___ Verbally Inappropriate	___ Self-injury	

Describe problem behavior: _____

Provide more detail about the problem behavior(s):

What does the problem behavior(s) look like?

How often does the problem behavior(s) occur?

How long does the problem behavior(s) last when it does occur?

What is the intensity/level of danger of the problem behavior(s)?

Identifying Routines: Where, When and With Whom Problem Behaviors are Most Likely.

Schedule (Times)	Activity	Likelihood of Problem Behavior	Specific Problem Behavior
		Low 1 2 3 4 5 6 High	
		1 2 3 4 5 6	
		1 2 3 4 5 6	
		1 2 3 4 5 6	
		1 2 3 4 5 6	
		1 2 3 4 5 6	
		1 2 3 4 5 6	
		1 2 3 4 5 6	
		1 2 3 4 5 6	
		1 2 3 4 5 6	

Appendix D: Implementation Fidelity Checklist

Implementation Fidelity Checklists

Please review if each of the following components were completed:

Protocol #1: Baseline

- ❑ Ensure signed consent
- ❑ Review the tool
- ❑ Review process of observations & time of observations
- ❑ Answer basic questions (do not review other components)

Time of Training: _____ minutes

Please review if each of the following components were completed:

Protocol #2: Behavior Basics

- ❑ Review human behavior & consequences
- ❑ Review 2 examples
- ❑ Review each consequence (AA, PA, Es, ATA)
- ❑ Importance of Function
- ❑ Connect importance of function and link to ISP-Fx
- ❑ Answer any questions

Time of Training: _____ minutes

Protocol #3: Performance Feedback

- ❑ Watch video #1
- ❑ Review correct answers and review incorrect responses
- ❑ Watch video #2
- ❑ Review correct answers and review incorrect responses
- ❑ Answer any questions

Time of Training: _____ minutes

Appendix E: IRB Approval Letter

USF
UNIVERSITY OF
SOUTH FLORIDA

RESEARCH INTEGRITY AND COMPLIANCE
Institutional Review Boards, FWA No. 00001669
12901 Bruce B. Downs Blvd., MDC035 • Tampa, FL 33612-4799
(813) 974-5638 • FAX (813) 974-7091

May 10, 2019

Casie Peet
Educational and Psychological Studies
Tampa, FL 33612

RE: **Expedited Approval for Initial Review**
IRB#: Pro00039544
Title: Facilitating Early Intervention Through Teacher Training in Brief Functional Behavior Assessment

Study Approval Period: 5/10/2019

Dear Ms. Peet:

On 5/10/2019, the Institutional Review Board (IRB) reviewed and **APPROVED** the above application and all documents contained within, including those outlined below. **Please note this study is approved under the 2018 version of 45 CFR 46 and you will be asked to confirm ongoing research annually in place of a full Continuing Review. Amendments and Reportable Events must still be submitted per USF HRPP policy.**

Approved Item(s):
Protocol Document(s):
Protocol Version 1_4.1.19

Consent/Assent Document(s)*:
Parent ICF_Version 1_4.1.19.docx.pdf
Teacher ICF_Version 1_4.1.19.docx.pdf

*Please use only the official IRB stamped informed consent/assent document(s) found under the "Attachments" tab. Please note, these consent/assent documents are valid until the consent document is amended and approved.

It was the determination of the IRB that your study qualified for expedited review which includes activities that: (1) present no more than minimal risk to human subjects, and (2) involve only procedures listed in one or more of the categories outlined below. The IRB may review research through the expedited review procedure authorized by 45 CFR 46.110. The research proposed in this study is categorized under the following expedited review category:

(7) Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

This research involving children as participants was approved under 45 CFR 46.404: Research not involving greater than minimal risk to children is presented.

Requirements for Assent and/or Permission by Parents or Guardians: 45 CFR 46.408 Permission of one parent is sufficient.

Assent is per application.

As the principal investigator of this study, it is your responsibility to conduct this study in accordance with IRB policies and procedures and as approved by the IRB. Any changes to the approved research must be submitted to the IRB via an Amendment for review and approval. Additionally, all unanticipated problems must be reported to the USF IRB within five (5) business days.

We appreciate your dedication to the ethical conduct of human subjects research at the University of South Florida and your continued commitment to human research protections. If you have any questions regarding this matter, please call 813-974-5638.

Sincerely,

Melissa Sloan, PhD, Vice Chairperson
USF Institutional Review Board

Appendix F: Parent Informed Consent Form

Study ID:Pro00039544 Date Approved: 5/10/2019

USF
UNIVERSITY OF
SOUTH FLORIDA

Parental Permission for a Child to Participate in Research
Information for parents to consider before allowing your child to take part in this research study

Title: Facilitating Early Intervention Through Teacher Training in Brief Functional Behavior Assessment

Pro # 00039544

Overview: We are asking you to allow your child to take part in a research study. The following information is being presented to help you and your child decide whether or not your child should participate in a research study. The sections in this Overview provide the basic information about the study. More detailed information is provided in the remainder of the document.

When we use the term "you" in this document, we are referring to your child.

Study Staff: This study is being led by Casie Peet who is a doctoral candidate and research assistant at the University of South Florida is. This person is called the Principal Investigator. She is being guided in this research by Dr. Nate von der Embse. Other approved research staff may act on behalf of the Principal Investigator.

Study Details: This study is being conducted at your child's school and is supported/sponsored by the National Association of School Psychologists. The purpose of the study is to find out what level of teacher training is necessary for your student's teacher to use a new assessment tool. Your child's teacher will receive different levels of training on how to measure student behavior (e.g. not following directions, talking to peers). To see how well your student's teacher can rate student behavior using the assessment tool, your child's teacher and the researcher will be observing your child. Your child will not be asked to do anything as a part of this research and your child will not be aware of their participation. The researcher will simply come into your child's classroom and observe your child for 10 minutes a day across 5-6 weeks.

Participants: You are being asked to take part because your teacher is participating and your child may display instances of disruptive behavior. This is to help your child's teacher understand more about your child and their behavior in the classroom in order to best meet your child's needs.

Voluntary Participation: Your participation is voluntary. You do not have to participate and may stop your participation at any time. There will be no penalties or loss of benefits or opportunities if you do not participate or decide to stop once you start. Your decision to participate or not to participate will not affect your student status, course grade, recommendations, or access to future courses or training opportunities.

Benefits, Compensation, and Risk: We do not know if you will receive any benefit from your participation. There is no cost to participate. You will not be compensated for your participation. This research is considered minimal risk. Minimal risk means that study risks are the same as the

risks you face in daily life.

Confidentiality: Even if we publish the findings from this study, we will keep your study information private and confidential. Anyone with the authority to look at your records must keep them confidential.

Study Procedures:

At each visit:

- The researcher will come into your child's classroom and observe your child for 10 minutes in their normal instructional time. Your child will not be aware that they are being observed. This will not impact their educational activities and they will not be removed from their classroom.

Total Number of Participants

About 12 individuals will take part in this study at USF. A total of 6 teachers and 6 students will participate in the study.

Alternatives / Voluntary Participation / Withdrawal

You do not have to participate in this research study.

You should only take part in this study if you want to volunteer. You should not feel that there is any pressure to take part in the study. You are free to participate in this research or withdraw at any time. There will be no penalty or loss of benefits you are entitled to receive if you stop taking part in this study. The decision to participate or not will not affect your student status, grade, or educational opportunities.

Benefits

We are unsure if you will receive any benefits by taking part in this research study. This research may help your child's teacher better understand your child and may inform educational decisions that could help your child succeed.

Risks or Discomfort

This research is considered to be minimal risk. That means that the risks associated with this study are the same as what you face every day. There are no known additional risks to those who take part in this study.

Compensation

You will receive no payment or other compensation for taking part in this study.

Costs

It will not cost you anything to take part in the study.

Privacy and Confidentiality

We will do our best to keep your records private and confidential. We cannot guarantee absolute confidentiality. Your personal information may be disclosed if required by law. Certain people may need to see your study records. These individuals include:

- The research team, including the Principal Investigator, study coordinator, and all other research staff.
- Certain government and university people who need to know more about the study. For example, individuals who provide oversight on this study may need to look at your records. This is done to make sure that we are doing the study in the right way. They also need to make sure that we are protecting your rights and your safety.
- Any agency of the federal, state, or local government that regulates this research. This includes: the Department of Health and Human Services (DHHS) and the Office for Human Research Protection (OHRP).
- The USF Institutional Review Board (IRB) and its related staff who have oversight responsibilities for this study, and staff in USF Research Integrity and Compliance.

Your information collected as part of the research, even if identifiers are removed, will NOT be used or distributed for future research studies.

We may publish what we learn from this study. If we do, we will not include your name. We will not publish anything that would let people know who you are.

Data collected for this research will be stored at the Department of Educational and Psychological Studies, located at the University of South Florida in the United States.

You can get the answers to your questions, concerns, or complaints.

If you have any questions, concerns or complaints about this study, call Casie Peet at (720) 987-4932. If you have questions about your rights, complaints, or issues as a person taking part in this study, call the USF IRB at (813) 974-5638 or contact by email at RSCH-IRB@usf.edu.

Consent for My Child to Participate in this Research Study

I freely give my permission to let my child take part in this study. I understand that by signing this form I am agreeing to let my child take part in research. I have received a signed copy of this form to take with me.

_____ _____

Signature of **Parent** of Child Taking Part in Study Date

_____

Printed Name of **Parent** of Child Taking Part in Study

_____
Printed Name of the **Child** Taking Part in Study

Statement of Person Obtaining Informed Consent and Research Authorization

I have carefully explained to the person taking part in the study what he or she can expect from their participation. I confirm that this research participant speaks the language that was used to explain this research and is receiving an informed consent form in their primary language. This research participant has provided legally effective informed consent.

_____ _____
Signature of Person Obtaining Informed Consent Date

_____
Printed Name of Person Obtaining Informed Consent

Appendix G: Teacher Informed Consent

USF
UNIVERSITY OF
SOUTH FLORIDA

Informed Consent to Participate in Research Involving Minimal Risk
Information to Consider Before Taking Part in this Research Study

Title: Facilitating Early Intervention Through Teacher Training in Brief Functional Behavior Assessment

Pro # 00039544

Overview: You are being asked to take part in a research study. The information in this document should help you to decide if you would like to participate. The sections in this Overview provide the basic information about the study. More detailed information is provided in the remainder of the document.

Study Staff: This study is being led by Casie Peet who is a doctoral candidate and research assistant at the University of South Florida is. This person is called the Principal Investigator. She is being guided in this research by Dr. Nate von der Embse. Other approved research staff may act on behalf of the Principal Investigator.

Study Details: This study is being conducted at your school and is supported/sponsored by the National Association of School Psychologists. The purpose of the study is to find out what level of teacher training is necessary for you to accurately use a new assessment tool. You will receive different levels of training on how to measure student behavior. The study will last approximately 6 weeks. You will receive three trainings that will last less than 15 minutes each and will take place in your classroom during one of your planning periods. Following each training, the principal investigator will come into your classroom to conduct a 10-minute observation on one of your students each day.

Participants: You are being asked to take part because you are a preschool teacher.

Voluntary Participation: Your participation is voluntary. You do not have to participate and may stop your participation at any time. There will be no penalties or loss of benefits or opportunities if you do not participate or decide to stop once you start. Your decision to participate or not to participate will not affect your job status, employment record, employee evaluations, or advancement opportunities.

Benefits, Compensation, and Risk: We do not know if you will receive any benefit from your participation. There is no cost to participate. You will be compensated up to $100 for your participation. This research is considered minimal risk. Minimal risk means that study risks are the same as the risks you face in daily life.

Confidentiality: Even if we publish the findings from this study, we will keep your study

information private and confidential. Anyone with the authority to look at your records must keep them confidential.

Study Procedures:

Trainings:

- You will have three separate training time spread across the 6 weeks. The first training will last 5-10 minutes. The second training will last 10-15 minutes. The final training will last 5-10 minutes. Each of these will take place in your classroom during one of your planning periods (e.g. before school, after school, lunch time). You will be trained by the principal investigator on how to use a new observation tool.

Observations:

- Each day, the researcher will come to your classroom to conduct a 10-minute observation on a student in your classroom with disruptive behavior. This will be the same student each day. We will acquire parental consent before beginning the study in order to observe this student. While the researcher is observing the student, you will simultaneously be observing the student (while continuing instructional practices) using the observational tool you were trained to use.

Total Number of Participants

About 12 individuals will take part in this study at USF. A total of 6 teachers and 6 students will participate in the study.

Alternatives / Voluntary Participation / Withdrawal

You do not have to participate in this research study.

You should only take part in this study if you want to volunteer. You should not feel that there is any pressure to take part in the study. You are free to participate in this research or withdraw at any time. There will be no penalty or loss of benefits you are entitled to receive if you stop taking part in this study. The decision to participate or not will not affect your student status, grade, or educational opportunities.

Benefits

We are unsure if you will receive any benefits by taking part in this research study. The training and use of the new measurement tool may help you to learn additional information about your students and help with educational and behavioral programming.

Risks or Discomfort

This research is considered to be minimal risk. That means that the risks associated with this study are the same as what you face every day. There are no known additional risks to those who take part in this study.

Compensation

You will be compensated $100 if you complete all the scheduled study visits in two separate payments of $50. If you withdraw for any reason from the study before completion you will be compensated $5 for each study visit you complete, up to $100.

It will not cost you anything to take part in the study.

Privacy and Confidentiality

We will do our best to keep your records private and confidential. We cannot guarantee absolute confidentiality. Your personal information may be disclosed if required by law. Certain people may need to see your study records. These individuals include:

- The research team, including the Principal Investigator, study coordinator, and all other research staff.
- Certain government and university people who need to know more about the study. For example, individuals who provide oversight on this study may need to look at your records. This is done to make sure that we are doing the study in the right way. They also need to make sure that we are protecting your rights and your safety.
- Any agency of the federal, state, or local government that regulates this research. This includes: the Department of Health and Human Services (DHHS) and the Office for Human Research Protection (OHRP).
- The USF Institutional Review Board (IRB) and its related staff who have oversight responsibilities for this study, and staff in USF Research Integrity and Compliance.

Your information collected as part of the research, even if identifiers are removed, will NOT be used or distributed for future research studies.

We may publish what we learn from this study. If we do, we will not include your name. We will not publish anything that would let people know who you are.

Data collected for this research will be stored at the Department of Educational and Psychological Studies, located at the University of South Florida in the United States.

You can get the answers to your questions, concerns, or complaints.

If you have any questions, concerns or complaints about this study, call Casie Peet at (720) 987-4932. If you have questions about your rights, complaints, or issues as a person taking part in this study, call the USF IRB at (813) 974-5638 or contact by email at RSCH-IRB@usf.edu.

Consent to Take Part in Research

I freely give my consent to take part in this study. I understand that by signing this form I am agreeing to take part in research. I have received a copy of this form to take with me.

_____ _____

Signature of Person Taking Part in Study Date

_____

Printed Name of Person Taking Part in Study

Statement of Person Obtaining Informed Consent and Research Authorization

I have carefully explained to the person taking part in the study what he or she can expect from their participation. I confirm that this research participant speaks the language that was used to explain this research and is receiving an informed consent form in their primary language. This research participant has provided legally effective informed consent.

_____ _____

Signature of Person Obtaining Informed Consent Date

_____

Printed Name of Person Obtaining Informed Consent