

2013

Outcomes of Program-wide Positive Behavior Supports in Early Childhood Settings

Robin Drogan
Lehigh University

Follow this and additional works at: <http://preserve.lehigh.edu/etd>

 Part of the [Special Education and Teaching Commons](#)

Recommended Citation

Drogon, Robin, "Outcomes of Program-wide Positive Behavior Supports in Early Childhood Settings" (2013). *Theses and Dissertations*. Paper 1476.

This Dissertation is brought to you for free and open access by Lehigh Preserve. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Lehigh Preserve. For more information, please contact preserve@lehigh.edu.

Outcomes of Program-wide Positive Behavior Supports
in Early Childhood Settings
By Robin Drogan

Presented to the Graduate and Research Committee
of Lehigh University
In Candidacy for the Degree of
Doctor of Philosophy
In
Special Education

Lehigh University

2013

Copyright by Robin Drogan

2013

Robin Ratner Drogan

Approved and recommended for acceptance as a dissertation in partial fulfillment of the requirements of Doctor of Philosophy.

Date

Dissertation Director
Lee Kern, Ph.D.
Professor of Special Education

Accepted Date

Committee Members:

George DuPaul, Ph.D.
Professor of School Psychology

Robin Hojnoski, Ph.D.
Associate Professor of School Psychology

Brenna Wood, Ph.D.
Assistant Professor of Special Education

Acknowledgements

My warmest thanks and sincerest gratitude go to my advisor, Dr. Lee Kern, for allowing me to grow and flourish under her guidance. I would like to thank my dissertation committee, Dr. George DuPaul, Dr. Robin Hojnoski, and Dr. Brenna Wood for their guidance and support throughout this process. In addition, I would like to thank the state grant facilitators, Sue Zeiders and Deb Hartman, and all of the early interventions, LeAnne, Suzanne, Holly, Amy, Gia, Heidi, Cindy, Janine, and Charlene, for their continuing commitment to this project.

My friends and family have been my ultimate supports throughout my graduate studies. Thank you to my wonderful husband, Bob, and my children, Jade and Connor. I needed them to keep me focused and to help me understand why to continue working towards my goals. I owe the biggest thanks to my parents. Their love has always been unconditional and I could not have done this without them. There are many people who have supported and challenged me with too many to mention. My friends at Lehigh have been the ultimate support, Amanda, Dolly, Emily, Talida, Cathy, and Clare. I hope these relationships will continue to grow and flourish. Lehigh will always have a special place in my heart and all of the professors, staff, and students here have made a difference in my life. I hope that the upcoming students will continue to build on the strong legacy of support and commitment that was available to me. My longtime friends, Heidi, Tami, Gary, Adrienne, Kate, Tricia, Pam, Mindy, and Tricia, have pushed me along and been there in times of need. I could not have made it through this journey without their love.

Table of Contents

Title Page.....	
Copyright Page.....	ii
Approval Page.....	iii
Acknowledgements	iv
Table of Contents.....	v
Abstract.....	1
Chapter 1: Statement of the Problem	2
Chapter 2: Review of the Literature.....	13
Chapter 3: Methodology.....	33
Chapter 4: Results	44
Chapter 5: Discussion	52
References	66
Tables and Figures.....	81
Vita.....	93

Abstract

There is growing concern over the high numbers of children who exhibit problem behavior in early childhood settings. Although initial data for a tiered framework of behavioral support, Program Wide Positive Behavior Intervention and Support (PWPBIS), has emerged and has shown promise in the early childhood literature, there are still a limited number of investigations on the topic. This study evaluated the effects of one region's PWPBIS implementation in six preschools with 21 teachers on challenging behavior in young children. A quasi-experimental research design evaluated the pre/post assessment results for student outcomes. In addition, the impact of teacher fidelity on student outcomes was evaluated using HLM. Results indicated that both the intervention and control group made gains across time. Fidelity of implementation did not affect student outcomes in this investigation. Rather, the teacher covariate of years of experience was statistically significant indicating that for each additional year of experience a teacher had, the intercept of the student post-test would decrease .186 units.

Chapter 1

Statement of the Problem

Challenging Behavior in Young Children

There are many children who enter early childhood settings without important social emotional skills necessary to interact productively, make friends, and solve problems. In the absence of these skills, children often display challenging behavior to meet their needs (Quesenberry, Hemmeter, & Ostrosky, 2011). In fact, teachers of young children have reported challenging behavior as the most problematic of the difficulties they face in preschool classrooms (Snell, Berlin, Voorhees, Stanton-Chapman, & Hadden, 2011). Early childhood providers report that addressing the needs of the children with the most significant social, emotional, and behavioral needs, in addition to providing a developmentally appropriate and supportive learning context for all children in the classroom, is a growing challenge (Hemmeter, Corso, & Cheatham, 2006).

There are many issues to consider when very young children exhibit problem behavior. These include the developmental age of the child, the cultural and familial expectations that make up the family value system, children's early experiences, and the presence of environmental supports to teach and promote appropriate behavior (Harry & Klinger, 2006). It is to be expected that young children will show some disruptive or oppositional behavior considering that they are just learning self-regulation, cooperation, and social adjustment skills (Greenspan & Weider, 2006; Kendziora, 2004). Within a developmental framework, as children learn effective strategies in communicative and social areas, behavior problems tend to diminish (Bambara & Kern, 2005). However, often behavior problems do not resolve as children age.

The term “challenging behavior” may be applicable for behaviors that occur persistently, significantly interfere with learning and social relationships, and are not responsive to strategies that consider the child’s developmental level, family and cultural style, and individual characteristics (Gleason, 2009).

Prevalence and Trajectory of Problem Behavior in Young Children

The prevalence estimates of challenging behavior in young children vary depending on methodology, instrumentation, and population. The estimated prevalence of young children with serious emotional/behavioral problems ranges from approximately 9.5%-14.2% (Brauner & Stephens, 2006). Research suggests that rates of problem behavior are higher for children of low-income families (Qi & Kaiser, 2003). In a review of 30 studies, Qui and Kaiser indicated that rates of problem behavior for young children from low-income families could be up to five times higher when compared to the general preschool population. Given these prevalence estimates, of the 24 million young children in the United States, 2.5 to 5.0 million young children may be at risk for exhibiting challenging behavior (Powell, Fixen, Dunlap, Smith, & Fox, 2007).

Preschool teachers report disruptive behavior problems to be the most important among the challenges faced in the classroom (Joseph & Strain, 2003). Challenging behavior often leads to social rejection by peers (Bierman, 2004). The developmental significance of social acceptance and rejection for young children is substantial. Problem behavior prevents children from forming social relationships (Fantuzzo, Bulotsky, McDermott, Mosca, & Lutz, 2003). Young children’s failure to develop positive peer relationships can predict poor social adjustment later in life and is associated with problems in academic performance (Buhs, Ladd, & Herald, 2006; Flook, Repetti, & Ullman, 2005; Ladd, 1990).

The effects of challenging behaviors are evidenced by high rates of expulsion from preschool settings. In fact, researchers have found that 10.7% of preschool teachers reported expelling one child in the last 12 months and 19.9% reported expelling more than one (Gilliam, 2005; Whitted, 2011). These expulsion rates are particularly alarming because the presence of significant and persistent challenging behavior at a young age leads to academic and social failure later in life (Dunlap et al., 2006; Kaiser, 2007; Webster-Stratton & Hammond, 1997). Research has shown that teacher ratings of children exhibiting poor learning related skills (attention, working memory, inhibitory control, and social emotional competence) in kindergarten predicted lower performance on reading and math measures in 6th grade as compared with higher rated peers (McClelland, Acock, & Morrison, 2006). In addition, children who engage in behavior problems early on have an increased chance of showing more serious behavioral concerns, have poor performance academically in later grades, and have frequently been retained in early school years (Qi & Kaiser, 2003; Raver & Knitzer, 2002).

Social and Emotional Development of Young Children

It is critical to meet the social emotional needs of young children. Although a variety of definitions exist, collectively the research suggests that social and emotional development and competence are clearly connected and represent distinct yet overlapping areas of behavioral functioning (Squires & Bricker, 2007). Social development refers to the behaviors, attitudes, and affect of children that are acquired during the formation of relationships with adults and peers (Ashiabi, 2007; Wittmer, Doll, & Strain, 1996). Emotions and emotional regulation play a major role in social development because the social interactions of children are often relational, involving the exchange between the child and event and the meaning that is attributed to the event (Halberstadt, Denham, & Dunsmore, 2002; Squires & Bricker, 2007). Children

demonstrate “emotional competence” when they have developed a range of emotional responses or expressions that they can manage internally to produce desired outcomes and satisfy both themselves and the expectations in a specific environment (Squires & Bricker, 2007; Wittmer et al., 1996).

In order for children to develop social and emotional competence, the goal is to teach them the skills necessary for success in life and to reduce the likelihood that they will engage in challenging behavior (Squires & Bricker, 2007). In a review of the literature examining pre-school children from low-income backgrounds with challenging behavior, Qi and Kaiser (2003) found that there was a significant relationship between problem behavior and social skill deficits. Some of the essential skills for social emotional learning are self-awareness (e.g., self concept), self-management (e.g., emotional, cognitive, and behavioral regulation), social awareness (e.g. identify and label emotions, perspective taking), relationship skills, and responsible decision making (e.g. moral values, social cognitive skills) (Zins, Bloodworth, Weissberg, & Walberg, 2007). Research has demonstrated that children who are considered socially competent exhibit the appropriate social skills and engage in fewer behavior problems. For example, Harden et al. (2000) found teacher rated social competence to be negatively associated with three types of behavior (hostile-aggressive, anxious-fearful, and hyperactive-distraction). When young children display social competence, skills have been built over time and in turn these children demonstrate successful interactions with peers and adults including early development of friendships (Brown & Dunn, 2008). Thus, effective strategies to teach pro-social behaviors to young children are needed.

Intervention to Address Challenging Behavior

A typical response to problem behavior in schools has been the use of punishment and exclusion (Skiba & Peterson, 1999). Yet punitive strategies have not proven sufficient in decreasing disruption or violence nor have policies (e.g., Zero Tolerance) resulted in improved behavior or school safety (Skiba, 2002; Skiba & Peterson, 2000). Research indicates that schools overemphasize the use of punishment and that inconsistencies exist with the delivery of punishment (Skiba, Horner, Chung, Rausch, May, & Tobin, 2011). For example, children from African American and Latino families are 2.19 times as likely to receive office disciplinary referrals (ODRs) in elementary school and more likely to receive suspension and expulsion for the same offenses as their Caucasian peers (Skiba et al., 2011). In fact, the use of punitive strategies to address problem behavior, without the use of positive support, results in an increase of a host of behavioral concerns including truancy, violence, and aggression (Guess, Helmstetter, Turnbull, & Knowlton, 1987; Mayer 1995). Rather, the use of evidence-based instructional practices that support the development of positive pro-social behavior in children has been emphasized (Center on the Social Emotional Foundations of Early Learning (CSEFEL), 2008; Dunlap et al., 2006; Powell et al., 2007; Smith & Fox, 2003).

Positive behavior support (PBS) is one approach that has gained increased attention in recent years due to its focus on teaching and reinforcing appropriate behavior. Further, PBS has been applied to children with challenging behavior with positive outcomes (Conroy, Dunlap, Clarke, & Alter, 2005). Developed originally as an alternative to aversive interventions used with individuals with significant disabilities, PBS has been expanded to a wide range of students and contexts (Horner, Sugai, & Anderson, 2010). Recent applications include school-wide positive behavior support (SWPBS). The rationale for SWPBS lies in the application of a

continuum of behavioral supports designed to meet all student's needs (Horner, Todd, Lewis-Palmer, Irvin, Sugai, & Boland, 2004; Sugai et al., 2004).

The SWPBS model utilizes a tiered approach to intervention. The model is implemented within three tiers that include intervention practices to meet all students' needs and organizational systems to sustain those practices (Horner et al., 2010). Each tier is comprised of specific core features that guide implementation. At the primary level of prevention, the whole school is the target for service delivery. Supports at the primary level include: (a) establishing a team comprised of representation from administration, school staff, and parents; (b) clearly defining three to five school-wide behavioral rules and expectations; (c) explicitly teaching and reinforcing a set of appropriate behaviors; (d) comprehensively implementing a school-wide positive reinforcement system; and (e) continually monitoring and evaluating progress through a team process (Lewis & Sugai, 1999; Sugai et al., 2000). At the secondary level, "selected" interventions provide behavioral or academic support to those who do not respond to universal supports (Walker et al., 1996). This generally consists of small group instruction or strategies that increase student monitoring, such as Check In Check Out (Hawken, MacLeod, & Rawlings, 2007; Simonsen, Myers, & Briere III, 2011). At the tertiary level, teams gather information, develop hypothesis, and implement function-based individualized intervention plans (Nahgahwon, Umbreit, Liaupsin, & Turton, 2010).

Recently, PBS has become more prevalent in preschool settings (Fox, Jack, & Broyles, 2005) with greater attention focused on the social emotional competence of young children. This shift has occurred for several reasons. First, it is optimal to intervene with children displaying challenging behavior at a young age (Stormont, Lewis, & Smith, 2005). Second, the relationship between the social emotional competence of young children and future academic

success has been established (McClelland, Cameron, Connor, Farris, Jewkes, & Morrison, 2007). Therefore, efforts are needed to increase the likelihood that young children are on a positive school trajectory and to avoid the onset of chronic negative behavioral and academic patterns. Further, support for social development is inherent in preschool programming (e.g., friendship skills, sharing, building relationships, feelings and emotions) and therefore aligns well with the use of positive behavior supports (Stormont et al., 2005). One way to meet the needs of more young children is to put into place a universal prevention based systems and provide early support for social and emotional development for all children, based on the idea that all children at a young age can benefit from guidance and instruction to learn how to express emotions, play cooperatively, and use problem solving skills (Joseph & Strain, 2003; Smith, Lewis, & Stormont, 2011).

The *Teaching Pyramid* represents one application of tiered positive behavior supports for young children. The framework provides guidance to early childhood educators regarding the use of developmentally appropriate evidence-based social emotional practices. There are differences in the application of positive behavior support in preschool that account for the unique variations that are developmentally necessary for young children. Some variations in the application for young children include: (a) two distinct parts to tier 1; (b) the distinction of program-wide supports rather than school-wide; and (c) the use of fewer behavioral expectations.

The research to date in the area of early childhood and the use of positive behavior support has been mostly descriptive in nature with a few studies evaluating the effectiveness of tiered supports. Only five studies were found evaluating the impact of program-wide Positive Behavior Intervention and Supports (PWPBIS) in early childhood settings. Results of these studies collectively demonstrated decreases in challenging behavior for target children (Benedict,

Horner, & Squire, 2007; Smith et al., 2011; Stormont, Smith, & Lewis, 2007) as well as children within PWPBIS programs (Muscott, Pomerleau, & Dupuis, 2009; Muscott, Pomerleau, & Szczesiul, 2009). Although four studies indicated a change in level of student problem behavior there were limitations in the studies. Only three of the five studies evaluated the effectiveness of PWPBIS (Benedict et al., 2007; Muscott, Pomerleau, & Dupuis, 2009; Muscott, Pomerleau, & Szczesiul, 2009) while two studies evaluated specific universal strategies with young children (Smith et al., 2011; Stormont et al., 2007).

For example, Smith et al. (2011) used a multiple-baseline design across teachers to investigate use of pre-corrective statements and behavior-specific praise as part of a prevention-based approach with young children identified as having high rates of challenging behavior. The teachers were asked to participate based on their low rates of praise and high rates of reprimand. Each teacher selected students 4-5 years of age who displayed low levels of appropriate social skills and high rates of problem behavior. The teachers received individual consultation in how to use the two targeted universal interventions and feedback regarding their use of the intervention. Results of this investigation demonstrated a functional relationship between teachers use of two universal strategies on student problem behavior across three target students. This study demonstrated the importance of implementation of universal supports for all children and the importance of using these supports first for children displaying high rates of challenging behavior (Stormont et al., 2007).

In a large scale evaluation of PWPBS in New Hampshire, two of nine early childhood programs reported data for implementation but the study lacked comparison to a non-intervention program (Muscott, Pomerleau, & Szczesiul, 2009). With the increased adoption of PWPBIS in preschools, more research is warranted to determine its effectiveness.

Treatment Integrity

Treatment integrity is commonly defined as the implementation of an intervention as intended in order to make valid conclusions regarding treatment outcomes (Lane, Bocian, MacMillan & Gresham, 2004). However, there is no consensus on the exact definition and broad characterizations have emerged to capture the complexity of the construct (Hagermoser Sanetti, and Kratochwill, 2009). The documentation of treatment integrity is essential to both the internal and external validity of intervention outcome research. For internal validity, treatment integrity data are necessary in order to make valid claims about the relationship of the independent and dependent variables (Moncher & Prinz, 1991). For external validity, adequate documentation of the application of the intervention is necessary for replication (Peterson, Homer, & Wonderlich, 1982).

In practical settings, assessing the degree to which interventions are implemented with integrity is valuable. First, research suggests that even with high levels of training, teachers fail to implement interventions with accuracy (DiGennaro, Martens, & McIntyre, 2005). Second, there is some research to suggest that student problem behavior is negatively correlated with treatment integrity (DiGennaro, Martens, & Kleinmann, 2007; Wilder, Atwell, & Wine, 2006). Third, teachers need to be able to determine whether a student's resistance to treatment is a result of inaccurate intervention implementation or ineffective intervention (Gresham, 2009; Moncher & Prinz, 1991).

Despite the importance of establishing integrity of implementation, this science remains underdeveloped (Hagermoser Sanetti, & Kratochwill, 2009). In the past, treatment integrity has received limited attention in both educational settings and in the published literature (Cochrane & Laux, 2008; Hagermoser Sanetti & Kratochwill, 2009). To date, careful attention has been

given in the intervention literature related to design, training, identifying the target audience, selecting outcome variables, and assessing the accuracy of the data collection; yet the integrity of implementation has been frequently excluded (Lane et al., 2004). In a recent review of school-based empirical studies published between 1991-2005 in *Journal of Applied Behavior Analysis (JABA)*, McIntyre, Gresham, DiGennero, and Reed (2007) found low levels of treatment integrity assessed. From a total of 142 articles included in the review, approximately 30% ($n=46$) of the studies reported the monitoring or documenting the intervention implementation. A review of 72 studies published in *Journal of Positive Behavior Interventions (JPBI)* between 1999-2009 Sanetti, Dobey, & Gritter (2012) found slightly higher assessment of treatment integrity, with 41.7% ($n=30$) of the studies monitored or assessed treatment integrity.

Recently, a resurgence emphasizing the need for practitioners and researchers in education to address treatment integrity has emerged. Many factors have collectively contributed to this renewed focus including recent federal legislation and the emphasis on evidence-based intervention (Hagermoser Sanetti & Kratochwill, 2009; Lane et al., 2004). There has been an increase in the rate of reporting treatment integrity data in the current literature, although the percentage remains at less than half of intervention studies. Therefore, more research is warranted investigating the effects of teacher fidelity on child outcomes.

Summary

In summary, limited research has been conducted to evaluate the effectiveness of PBIS early childhood settings. More information is needed on tiered implementation of PWPBIS for young children. Further, there is a need to investigate child outcomes in a tiered intervention program compared to a control group. Additional data are needed to determine the impact of teacher fidelity on student outcomes. Therefore, there is a need to further evaluate the

implementation of PWPBIS across multiple early childhood settings as compared to typical classroom conditions. The current study will extend the existing literature base by further evaluating the effects of PWPBIS on child outcomes using intervention and comparison schools. In addition, the relationship between teacher implementation and child outcomes will be examined.

Research Questions

Research Question 1: Do post-test outcomes of students from preschools that participated in PWPBIS across an academic year show significantly greater growth from pre-test to post-test on a standardized measure of social emotional skills than children in preschool classrooms that did not participate in PWPBIS?

It was hypothesized that there would be a statistically significant increase from pre-test to post-test in social emotional skill outcomes of children whose preschools participated in PWPBIS as compared to those who did not participate in PWPBIS.

Research Question 2: Do teacher variables (fidelity of implementation, years of experience, education level) predict post-test child social emotional outcomes while controlling for pre-test social emotional status?

It was hypothesized that teacher variables (fidelity of implementation, years of experience, education level) would predict higher post-test child outcomes while controlling for pre-test social emotional status.

Chapter 2

Literature Review

Research Supporting Positive Behavior Intervention and Support (PBIS)

Researchers and educators have recognized that intervention and prevention to address the social emotional needs of young children during preschool years has a long-term positive impact on the social outcomes of children (Center on the Social Emotional Foundation of Early Learning [CSEFEL]; Kendziora, 2004). This has resulted in an increase in the adoption of Positive Behavior Interventions and Supports (PBIS) in early childhood settings. Several descriptive studies have addressed how program-wide Positive Behavior Interventions and Supports (PWPBIS) can be implemented to support children in early childhood settings (Fox & Little, 2001; Stormont, Lewis, & Beckner, 2005). Preliminary investigations of the use of universal prevention to address challenging behavior in young children demonstrated promising results (Stormont et al., 2007; Smith et al., 2011). To date, however, no empirical studies have been conducted on the effectiveness of the implementation of PWPBIS in early childhood settings.

Evaluations of PBIS with school-aged children provide evidence for the potential benefits of school-wide prevention and intervention for young children. School-wide Positive Behavior Intervention and Support (SWPBIS) are being implemented in over 18,000 schools across the United States in various stages of adoption and in several other countries around the world (Horner et al., 2010). Further, there is a growing number of studies empirically evaluating the efficacy of SWPBIS (Bradshaw, Mitchell, & Leaf, 2010; Bradshaw, Reinke, Brown, Bevans, & Leaf, 2008; Horner et al., 2009). Although these large-scale studies support the effectiveness of SWPBIS, some skepticism remains as to whether SWPBIS is considered an evidence-based

practice. Therefore, researchers have begun to explore whether there is sufficient data to support SWPBIS as an evidence-based practice. For instance, Lane, Robertson, and Graham-Bailey (2006) investigated the methodological strengths and limitations of the SWPBIS literature between 1990-2005. The quality of the research was examined in relation to school characteristics, intervention components, research design, reliability, validity measurement, and intervention outcomes. Fourteen studies were identified implementing SWPBIS. Although many of the 14 studies produced favorable outcomes, Lane et al. (2006) found a number of limitations in methodology. The studies reviewed were primarily descriptive in nature (n=10, 71%), lacked sufficient detail in describing the intervention, and lacked reliability and validity information for the dependent measure (n=9, 64%). Therefore, Lane et al. concluded that due to methodological limitations, it was difficult to draw accurate conclusions about the intervention outcomes and described the findings overall as cautiously optimistic.

In addition, Horner et al. (2010) analyzed the body of SWPBIS literature published between the years of 2000- 2009. They developed five critical methodological features to determine whether SWPBIS has sufficient support to be considered an evidence-based practice. Forty-six articles were reviewed and represented a sample of the current peer-reviewed literature that directly addressed SWPBIS implementation and effectiveness. Horner and colleagues states that the articles selected were not intended to be a comprehensive review of all of the SWPBIS literature but a sample of the current research that directly addresses implementation and effectiveness. Within this sample, 20 articles focused on primary prevention, 13 focused on secondary prevention, and 13 focused on tertiary prevention. The criteria, originally designed for consideration of individual studies, included the following: (a) the practice and participants were defined with operational precision; (b) the research employed valid and reliable measures;

(c) the research was grounded in rigorous design; (d) the research documented experimental effects without iatrogenic outcomes; and (e) the research documented outcome effects (Horner et al.). Horner et al. (2009) discussed that these criteria may provide a useful organization for understanding if SWPBIS has sufficient empirical support to warrant consideration as an evidence-based practice in the three tiers. Although the criteria for examining the evidence base for any practice are typically applied to individual studies, here the criteria were applied to the research as a whole to be viewed together as a collective body. In their analysis, Horner et al. (2010) reported the practices associated with SWPBIS implementation, as demonstrated by the literature as a whole, met the five criteria and provided support in the following ways. They stated first that the studies provided a set of procedures that can be defined with replicable precision. In addition, they argued that a mix of standardized assessment measures and direct observation were used, including office discipline referrals and measurement of SWPBIS fidelity (e.g., Individual Student Systems Evaluation Tool (ISSET); Anderson et al., 2008; School-Wide Evaluation Tool (SET); Sugai, Lewis-Palmer, Todd, & Horner, 2001) and that together these measures assessed implementation with valid and reliable indices of fidelity and effect. Further, Horner et al., reported strong experimental effects with a mix of randomized control design studies (e.g., Bradshaw, Koth, Bevans, Ialongo, & Leaf, 2008; Bradshaw, Koth, Thornton, & Leaf, 2009; Horner, et al., 2009) combined with single subject designs (with a majority of the studies investigating Tier II and Tier III practices) (e.g., Beard-Jordan & Sugai, 2004; Carr et al., 1999; Sprague & Perkins, 2009). They argued that together the results from these randomized control trials, quasi-experimental studies, and systematic evaluations demonstrated a pattern of implementation fidelity associated with improved student behavior (Horner et al., 2010).

In response to discrepancy in the conclusions of Lane et al. (2006) and Horner et al. (2010), Chitiyo, May, and Chitiyo (2012) attempted to extend the earlier work of Horner et al. (2010) by applying the five criteria method to individual studies, rather than to the body of SWPBIS research as a whole. Studies were reviewed that were published in peer reviewed journals between 1990 and July 2011. The initial criteria included experimental studies that reported universal (tier 1) level behavioral interventions and also included student outcome data. Thirty-four articles met the initial criteria; however, 24 descriptive, non-experimental studies were excluded, resulting in a total of 10 studies included for analysis. Results of this investigation indicated that only three of the ten studies were grounded in rigorous designs. Although most of the studies reported implementation fidelity, only two of the ten studies reported high fidelity. Most importantly, in contrast with the findings of Horner et al. (2010) and more consistent with the Lane et al. (2006) findings, only two of the ten studies met all five of the criteria for high quality evidence-based practice, suggesting that further inquiry into the efficacy of SWPBIS is warranted.

Outcome of Randomized Controlled Trials

A few relatively recent randomized controlled trials have evaluated the effectiveness of SWPBIS at the universal level. Bradshaw et al. (2010) reported data to support reductions in problem behavior through implementation of SWPBIS in a randomized control effectiveness trial with 37 elementary schools. Schools were matched on specific demographic variables and then randomly assigned to intervention and control groups. In this efficacy trial, all training and support for universal prevention were coordinated and led by the SWPBIS Maryland State Leadership Team. The schools in the comparison condition did not receive training and were not eligible to register for the trainings provided by the state. The Effective Behavior Support

Survey (EBS; Sugai, Todd, & Horner, 2000) and SET (Sugai et al., 2001) were used to measure implementation fidelity. Implementation fidelity data indicated that schools trained in SWPBIS implemented with high fidelity as rated by staff through self-report (e.g., EBS) and through assessments conducted by outside evaluators who were kept unaware of the primary purpose of the study. Analysis of the EBS data indicated a statistically significant effect of training on the percentage of staff rating the school as having school-wide systems in place ($p < .001$), including non-classroom settings ($p < .001$), classroom settings ($p = .01$), and individualized students systems ($p = .001$). In addition, the SET data, which determined the degree to which the school had implemented each of seven critical features of the SWPBIS model, suggested a significant intervention effect for the overall SET score ($p = .001$). Six of the seven subscales demonstrated significant intervention effects including the following: define expectations ($p = .001$), teach behavioral expectations ($p = .001$), reward system ($p = .001$), monitoring and decision making ($p = .001$), management ($p < .001$) and district support ($p = .014$). The seventh subscale, behavioral violations ($p = .073$), did not reach statistical significance when all four years of data in the repeated measure were included. Office Disciplinary Referrals (ODR), suspension rates, and achievement scores were collected to assess student outcomes. ODR data were collected through the School-Wide Information System (SWIS; May et al., 2003) data collection procedure and did not include data for the baseline year. In addition, ODR data for intervention schools could not be compared to comparison schools because the comparison schools did not use SWIS. Therefore, ODR data from SWPBIS schools only were analyzed using repeated measures across the four post-training years to determine if there were significant differences. The percentage of students with major or minor ODRs decreased significantly over the course of intervention as well as the number of students with major and minor ODR events per student. The data revealed

that reduction of problem behavior was evidenced by fewer ODRs associated with problem behavior and a reduced proportion of students receiving out-of-school suspensions during implementation. A limitation, however, was the absence of both data from the non-SWPBS school and comparative baseline data from SWPBS schools.

Horner et al. (2009) found similar results with low levels of ODRs after universal implementation in a randomized control trial of 63 elementary schools across Illinois and Hawaii. Typical state personnel conducted all direct intervention with the participating schools. In this investigation, the SET was used to measure implementation fidelity, the School Safety Survey (SSS; Sprague, Colvin, & Irvin 1996) was used to rate perceived school safety, and the level of problem behavior was measured with report of ODR using SWIS. In addition, students were measured on the state reading assessment annually. Statistically significant increases were seen in the SET total score, demonstrating that personnel could use the SWPBIS training materials and implement the SWPBIS practices. Further, SWPBIS schools were perceived as safer environments with statistically significant differences on the SSS between SWPBIS and control schools. The results of ODR data indicated comparatively low rates of ODRs; however, the decrease was not necessarily associated with SWPBIS implementation because there were no baseline ODR data. Results of academic outcomes indicated no significant increases. Therefore, the association between SWPBIS and increased reading performance, measured by standardized assessments, should be considered with caution.

Both Horner et al. (2009) and Bradshaw et al. (2010) examined the interaction effects of behavioral as well as academic outcomes. To measure the impact of SWPBIS on academic outcomes, the percentage of students meeting or exceeding proficiency on the state academic performance standard was assessed annually in each study. Consistent findings in the area of

school achievement were demonstrated with non-significant results, although promising trends were noted. Specifically, Horner et al. reported that the Time X Condition effect was not statistically significant for the proportion of students that met or exceeded the state reading standard. Bradshaw et al. (2010) reported that although greater gains on math and reading scores were seen by students participating in SWPBIS schools as compared to the comparison schools, non-significant results were obtained. These results suggest an interaction between discipline and academic performance may exist but they do not show that implementation of SWPBIS leads to increased reading achievement or mastery of literacy skills.

Although SWPBIS is seen as a promising approach, there are a number of limitations in the research base suggesting that more inquiry is necessary. First, only half of the studies reviewed by Chitiyo et al. (2012) included adequate specificity in the description of the participants (both teachers and students) and explanation of the practice that would allow for replication. The failure to report student and teacher characteristics limits the ability to generalize the findings and to identify factors that predict student outcomes (Lane et al., 2006). Bradshaw et al. (2010) suggested that in the efficacy trial of SWPBIS in Maryland, the researchers did not have detailed information on the training or planning activities conducted by the state, even though a treatment integrity measure documented the activities that occurred. Second, there are concerns regarding the sole use of ODRs in measuring the reduction of problem behavior. The major problem is that ODRs are a subjective measure of behavior. The variability due to the number of people involved in the referral process and the potential complexity of the interactions between the players can be problematic for ensuring consistent outcomes (Irvin, Tobin, Sprague, Sugai, & Vincent, 2004). For example, Irvin et al. found that although ODRs were associated with problem behavior, the specific behaviors that were reported

varied greatly. Although this remains a concern, many studies use a combination of outcome measures. Eight of ten studies reviewed by Chitiyo et al. (2012) included ODR measures along with additional measures such as, suspension, academic achievement, and direct observation. Suspension and academic achievement data, however, also are subjective and may reflect bias. Third, although most studies monitored and evaluated treatment fidelity, only two studies reported high fidelity (Chitiyo et al.). The fidelity of implementation helps to determine whether the outcome can be attributed to the intervention. Finally, SWPBIS has been evaluated across educational settings from preschool to high school. However, evaluation at the preschool and high school levels is limited, with only two studies at each level identified by Chitiyo et al. and one high-school study identified by Lane et al. (2006).

PBIS at the Preschool Level

Although the research base for PBIS is growing, particularly at the elementary level, few studies have been conducted in preschool settings. Five studies were found that evaluated child outcomes at the preschool level (Benedict et al., 2007; Muscot, Pomerleau, & Dupuis, 2009; Muscott, Pomerleau, & Szczaesiul, 2009; Smith et al., 2011; Stormont et al., 2007). Four of the studies demonstrated decreases in problem behavior; however, each used different instrumentation (Muscot, Pomerleau, & Dupuis, 2009; Muscott, Pomerleau, & Szczaesiul, 2009; Smith et al., 2011; Stormont et al., 2007).

Using a multiple baseline across four classrooms, Benedict et al. (2007) assessed the relationship between PWPBIS consultation, teacher implementation of PWPBIS, and child problem behavior. An initial meeting was held to review the Pre-SET results and create action plans with the lead teachers and teams of support staff. In subsequent meetings, the consultant observed in the individual classrooms, modeled for the teachers, and provided ongoing verbal

and written feedback. Three measures were used to assess PWPBIS implementation and child behavior: (1) Preschool-wide PBS Fidelity Checklist; (2) direct observation of problem behavior; and (3) social validity using the PBS Consultant Questionnaire. As a result of consultation, changes were seen in the implementation of PWPBIS features in each of the four classrooms. The most significant changes occurred in two implementation areas: behavior acknowledged; use of a systematic approach to responding to behavior; and management, including items related to the teams focus on addressing PBS goals throughout the school year. General rates of problem behavior were low across phases, with a mean of 2.76% (range 0-20%) therefore no discernible changes in level, trend, or variability were seen between baseline and post-consultation. A decrease in the mean percentage of intervals with problem behavior was seen in three of the four classrooms. Social validity scores indicated high acceptability with mean scores ranging from 4.50-5.89.

In another study, Muscott, Pomerleau, and Dupuis (2009) described a case study of the implementation of PWPBIS at the Visiting Nurses Association (VNA) Child Care and Family Resource Center in New Hampshire. The VNA site was recruited as a part of a statewide PBIS initiative that began the year prior and VNA was accepted into Cohort 2 along with two other Head Start Programs and 24 public schools (Muscott, Mann, Benjamin, Gately, Bell, & Muscott, 2004). Training and support for the VNA consisted of developing a leadership team, creating an assessment process, and engaging staff in technical assistance on the PBIS framework and primary prevention interventions. Three measures were used to assess treatment fidelity: (1) The Universal Team Checklist (UTC; Sugai, Horner, & Lewis-Palmer, 2000); (2) the EBS; and (3) the SET. Recordable incidences of challenging behavior were collected to assess student

outcomes. Behavior data were collected through the SWIS data collection system and did not include data for the baseline year.

Results were reported over three years of implementation during the school years 2005-2006 to 2007-2008. Treatment fidelity measures showed significant improvements over baseline, winter 2004 to spring 2006, with high fidelity. Follow-up SET data (year 2007-2008) indicated that VNA sustained adequate average fidelity after training and technical assistance ended. For challenging behavior, frequency and percentage of change data were reported across the three years. Between the 2005-2006 and the 2007-2008 school years, results indicated an 87% decrease in the top three challenging behaviors (aggression, defiance/disrespect, and abusive language). This study demonstrated an initial investigation of PWPBIS showing positive outcomes; however, there were several limitations. Although decreases were seen in problem behavior across years, the change was not necessarily attributable to the intervention since no baseline data were reported. In addition, student outcome measures were limited to frequency of problem behavior as reported by the teachers. It is unclear as to whether this measurement procedure captures objective and consistent data.

A third study included a large scale implementation of PWPBIS in early childhood education programs in New Hampshire. Muscott, Pomrleau, and Szcuzesiul (2009) reported child outcome data for two out of nine programs that implemented PWPBIS. The program sample included two programs from Cohort 2, including the site reported previously (Muscott, Pomerleau, & Dupuis, 2009). The first phase of implementation after initial recruitment was readiness. The universal leadership team was developed to oversee implementation and begin initial training. Training included six full days and included all teams. Adaptations were made to the PBIS framework with consideration to the differences between school-aged and early

childhood implementation. First, the terminology was established as program-wide rather than school-wide, and the use of the word challenging, rather than problem, was used to describe behavior. Second, an ODR form was adapted based on the early childhood programs' needs resulting in a behavior incident class form. The form included date, type of behavior, routine in which the behavior took place, the perceived motivation, and the teacher response. Third, changes were made to ensure the use of developmentally appropriate supports. Two to three program-wide expectations were chosen using words that were understandable to young children. In addition, a behavioral matrix of clearly defined expectations was developed in the context of daily routines (arrival, mealtime, center time, circle), rather than focusing on locations (classroom, hallway, bathroom, cafeteria). Results indicated that both sites had significantly fewer reportable incidences of problem behavior after PWPBIS implementation. A limitation, however, was that the study was descriptive and did not report baseline data for either site. The authors did not report the frequency and percentage of decrease for each year of implementation individually. The data were incomplete, with percentage of decrease for the first year of implementation reported for only one site and total number of behavioral incidents between the first and third years of implementation reported for both sites combined.

A fourth study used a single-subject evaluation of three teacher-student dyads (Smith et al., 2011). In this investigation, teachers were trained to use two targeted universal strategies, pre-corrective statements and behavior specific praise based on their low rates of praise and high rates of reprimands. The schools were in the second year of implementing program-wide PBIS. Three measures were used to assess teacher behavior and student behavior: (a) Social Skills Rating System-Teacher Form preschool version (SSRS-TF; Gresham & Elliott, 1990); (b) The Multi Option Observation System for Experimental Studies (MOOSES; Tapp, Wehby, & Ellis,

1995); and (c) a seven item social validity checklist. First, teachers selected possible students in their classrooms for assessment purposes based on observation and the Social Skills Rating System-Teacher Form preschool version (SSRS-TF; Gresham & Elliott, 1990). Children who displayed low levels of appropriate social behaviors and high rates of challenging behaviors were included. Subsequently, the Social Skills Rating System-Teacher Form (SSRS-TF; Gresham & Elliott, 1990) preschool version was used as a screening tool. The SSRS-TF includes two subscales in problem behavior (externalizing and internalizing) and three subscales in social skills (cooperation, assertion, and self-control). Children who scored at or greater than the 75th percentile on the SSRS-TF problem behavior subscale were selected to participate. In addition, children with SSRS-TF social skills subscale scores that were rated at or below the 25th percentile were desired although this was not a requirement. Student outcomes were measured using the SSRS-TF post-intervention. Additional outcome measures included direct observation of specific child problem behavior (frequency converted to rate per minute) and a social validity survey completed by the teacher. Direct observation of student behavior during a teacher-directed large group activity recorded the presence of on-task behavior and aggressive behavior exhibited by the target student. The social validity checklist was developed by the authors and was used to determine teacher perceptions of the intervention. A 5-point Likert scale was used with responses ranging from strongly disagree (1) to strongly agree (5).

As a result of the implementation of two universal behavioral supports, an overall improvement in child behavior was observed. The teacher of one student was unavailable to provide a post-score SSRS-TF rating. The SSRS-TF resulted in overall increases in social skills and decreases in problem behavior for the two children, for whom data were available, of three students. In addition, all three students showed increases in on-task behavior and decreases in

aggressive behavior, measured by direct observation of duration of on-task converted to percentage of time and frequency of aggressive behavior converted to rate per min. Scores on the social validity checklist ranged from 4-5 with only one of the seven items having a score of 4. Based on the content of the questionnaire, results of the social validity checklist suggested that teachers valued the intervention and perceived changes in the student's behavior.

In a fifth study, Stormont et al. (2007) investigated the relationship between teachers use of key universal features of PWPBIS on the rate of children's problem behavior. Two critical behavioral supports were identified as areas of improvement for these teachers, the use of specific praise and precorrection. Teachers 1 and 3 were teachers certified in education or a related field and Teacher 2 was a teaching assistant with prior experience in preschool classrooms. The participants were selected from a larger study which included data on teachers' use of praise and reprimands. The intervention took place in two classrooms, one teacher had her own classroom while the other teacher and the teaching assistant both implemented in different areas of the same classroom. Although this posed a potential threat to the integrity of the intervention, precautions were taken to avoid any potential interference. The researchers specifically told the teachers not to discuss the intervention and the small groups were arranged on opposite sides of the classroom which made it difficult to hear the implementation. Measures of teacher and student behavior included: (a) the Teacher Behavior Observation Form for observation of specific behavior praise, precorrection, and reprimand; (b) student behavioral observation including off-task, oppositional, disruptive, aggressive, and other types of externalizing behavior; and (c) a social validity survey. Student problem behavior data were collected using frequency counts within intervals on all children during the small group. Teachers were instructed to use pre-corrective statements to orient student to the lesson before

initiating the lesson and to increase rates of specific praise statements when students were following the behavioral expectations. Teacher training occurred during a 30-min meeting using a standard template of the content for consistency, including examples from small group activities observed during baseline, and providing practice with pre-corrective statements. During the intervention phase, the teachers received feedback on whether they had used the pre-correction strategy at the beginning of the activity and the specific number of praise statements they gave. Results indicated that problem behavior decreased for all students. However, problem behavior showed a decreasing trend during baseline prior to intervention implementation for the third teacher's class and teacher reprimand data were variable for all participants during baseline and thus it is difficult to say with certainty that these decreases were attributable to the intervention.

Treatment Fidelity Outcomes

There has been a great amount of empirical attention given to the construct of treatment integrity. Significant questions remain regarding how treatment integrity should be defined and how it should be addressed in research and practice. Although many conceptual models of treatment integrity have emerged, there is no consensus on what dimensions should be included in a comprehensive conceptualization (Hagermoser Sanetti & Kratochwill, 2009). A majority of the models proposed (Noell, 2008; Power et al., 2005; Waltz, Addis, Koerner, & Jacobson, 1993) have included at least one or more of the following dimensions: (a) content, (b) quality, (c) quantity, and (d) process (Hagermoser Sanetti & Kratochwill). A multidimensional construct explains the varying degrees of implementation (e.g., what was delivered, how well it was delivered, how much was provided, and how it was delivered). The following working definition of treatment integrity was proposed to incorporate a multidimensional

conceptualization: “Treatment integrity is the extent to which essential intervention components are delivered in a comprehensive and consistent manner by an interventionist trained to deliver the intervention” (Hagermoser Sanetti & Kratochwill, p. 448).

In addition, treatment integrity is complex because of the many moderators and mediators that can influence implementation. Hagermoser Sanetti and Kratochwill (2009) organized these variables around the four categories of external environment, organization, intervention, and interventionist. External environment refers to the level of support from stakeholders and policies. Organization refers to the resources available for training and implementation. Intervention refers to the complexity, feasibility, efficiency, and effectiveness. Interventionist refers to the perceptions, motivations, and skills of person implementing. Generally, as the complexity and time requirements increase, the level of treatment integrity decreases (Lane et al., 2004). Additional variables associated with the interventionist may include the amount or type of teacher training, teacher stress, or the teacher’s previous experience with behavioral interventions (Noell & Gansle, 2006).

Teacher qualifications, as variables associated with practices, have become important as policymakers and programs advocate for high standards to produce quality and effective programming. Many states are now requiring that lead early childhood teachers hold a Bachelor’s degree (NIEER; Barnett et al., 2005). Though, research suggests that teacher education level is not necessarily associated with improved classroom quality or child outcomes (Early et al., 2007; Pinata et al., 2005). It seems that teacher attributes may play a role in accounting for quality through a variety of paths (Pinata et al). For example, in a study of 238 pre-kindergarten classrooms that measured quality, Pinata et al. found that the teacher having a bachelor’s degree with specialized training in early childhood education had statistically higher

quality (in some areas) than the degree itself. More experience teaching young children and more child-centered (non-traditional) beliefs were also related to higher quality. These associations were small but were confirmed by Early et al. (2007). Results of a review of seven major studies were mixed related to the association between teachers' education and either classroom quality or student outcomes. The relationship between teachers' education and quality was evident only when simple analysis techniques were used but when more predictors were added, the effects diminished. These findings collectively suggest that classroom quality and child outcomes may be influenced by a host of components related to teachers in terms of training, experience, attitudes as well as more distal factors in the program ecology.

Treatment integrity may also be influenced by student variables. For examples, there may be differences in implementation when students are acquiring a new skill compared to when students are developing fluency or when an intervention is initially implemented with a student compared to an intervention in which the student has a history of exposure (Hagermoser Sanetti & Kratochwill; Noell, Gresham, & Gansle, 2002).

Few studies of child behavior have demonstrated the importance of measuring and reporting treatment fidelity and the effects on implementation outcomes. Northup, Fisher, Kahng, Harrel, and Kurtz (1997) evaluated the fidelity of implementation at various levels for a differential reinforcement plus time out procedure. Fidelity of implementation occurred at various levels with appropriate behavior being reinforced on 100% of occasions, 50% of occasions, or 25% of opportunities. The time out procedure, implemented for aberrant behaviors, followed the same variable pattern of implementation. The results of this investigation showed that the intervention effects were maintained, regardless of whether treatment integrity was 100% or 50%. These findings indicate the need for further investigation

to examine variability with respect to both implementation and floor levels necessary for behavior change.

In another study, Vollmer, Roane, Ringdahl, and Marcus (1999) examined differential reinforcement at varying integrity levels. In this study it was found that after the child had been exposed to DRA at 100% integrity, implementation at lower levels of integrity did not produce same outcomes. The results of a study by Wilder, Atwell, and Wine (2006) were consistent with that of Vollmer and colleagues. In an investigation of a three-step prompting procedure, Wilder and colleagues found that the implementation integrity has a large impact on the effects.

In a study of young children, Arkoosh, Derby, Wacker, Berg, McLaughlin, and Barretto (2007) evaluated a home-based early intervention package that was implemented by parents. Intervention was developed through functional analysis by the researchers. These data were derived from a prior study published by Derby, et al., 1997. Treatment fidelity was evaluated by observation of video sessions for the contingent presentation of a response (reinforcement or aversive) following the child's behavior (appropriate or inappropriate). The treatment integrity results generally supported the notion that higher treatment integrity levels lead to more positive results. The treatment integrity for positive responses was consistently higher. The treatment integrity for aversive responses was low for all students despite long-term behavior change. This may substantiate the premise that not all intervention components are equally important and that the amount of treatment may differ per component (Gresham, 2009, Hagermoser Sanetti & Kratochwill, 2009).

Ensuring that practices associated with positive child outcomes are being implemented with fidelity is equally important in the early childhood program-wide PBIS model. Thus, the assessment of teacher level of implementation of the model is a critical area of need. In general,

there are few treatment fidelity tools that are psychometrically sound and available to researchers (Sanetti et al., 2012) and this is true for assessing fidelity of the PWPBIS implementation. The investigations of PWPBIS in early childhood settings in New Hampshire assessed treatment fidelity (Muscott, Pomerleau, & Dupuis, 2009; Muscott, Pomerleau, & Szczesiul, 2009). Muscott, Pomerleau, and Szczesiul used the SET to assess implementation fidelity with four of the five cohorts and the Pre-School Evaluation Tool with the fifth cohort (Pre-SET; Horner, Benedict, & Todd, 2005). The Pre-SET had not been created at the onset of the study so could not be used with all cohorts. The researchers reported that implementation may be improved with the use of a tool that was specifically designed for early childhood settings.

The Pre-SET has since been adapted from the SET to address the unique context of early childhood settings implementing PWPBIS. The 30 items on the Pre-SET are organized into eight subscales that correspond to the features of PWPBIS at the universal tier of intervention. Administration of the Pre-SET should occur across an entire program to evaluate the implementation of PWPBIS and universal support although the Pre-SET can be used at the individual classroom as well as the program level as well. The PreSET and TPOT both use a consultant based process for gaining information on implementation. An outside observer conducts interviews and direct observations in order to complete the fidelity tools. Unlike the Pre-SET, the TPOT assesses all three of the levels of tiered support (universal, targeted, and tertiary). The Pre-SET has a similar scoring to the SET, using a three point Likert-scoring system, ranging from 0, not implemented, to 2, fully implemented. The Pre-SET scoring system utilizes the same scale as the SET that is used in many school aged programs implementing PBIS. The TPOT uses a dichotomous yes/no response for the first 7 items, a rating of 1-5 based on observation and interview for items 8-18, and a checklist for items 23-38. Initial

psychometric properties of the Pre-SET have been published and though the TPOT was used as a comparison for convergent construct validity, the psychometric properties of the TPOT have yet to be fully reported.

An initial analysis of the psychometric properties of the Pre-SET (Steed & Webb, 2012) demonstrated strong internal consistency, with an overall of alpha .91. Interobserver agreement on individual PreSET items averaged 95% (range= 68%-100%). The Pre-SET and the Teaching Pyramid Observation Tool for Preschool Classrooms (TPOT; Hemmeter, Fox, & Snyder, 2008), another implementation measure for preschool settings, were found to be moderately and positively correlated ($r=.33$), demonstrating that the Pre-SET has initial convergent construct validity with another tool used to assess PWPBIS. In addition, Paired *t*-tests compared the pre-implementation to post-implementation mean PreSET scores ($t = 10.49$ ($df = 28$), $p < .000$) suggesting that the Pre-SET is sensitive to implementation change (Steed & Webb, 2012).

In sum, treatment integrity is a complex and multidimensional construct. Overall, research suggests that higher levels of integrity results in better outcomes, (Noell 2008; Noell, Gresham, & Gansle, 2002; Wilder et al., 2006) although low levels of integrity are not necessarily related to poorer outcomes (Noell, 2008) and may have either a negative, neutral, or positive effect (Hagermoser Sanetti, & Kratochwill, 2009). In addition, all intervention components may not be equal and some components may actually be more critical than others (Noell, 2008). The majority of the studies reviewed have looked at fidelity of implementation when the intervention is introduced to individual students. In addition, teacher and student variables may impact fidelity (Noell, Gresham, & Gansle, 2002; Severson, Walker, Hope-Doolittle, Kratochwill, & Gresham, 2007). There is a need to further understand what level of integrity is necessary for specific treatments to produce beneficial outcomes (Gresham, 2009;

Noell et al., 2002). More specifically, there is a need to measure the effects of fidelity of implementation on student outcomes within the PBIS framework.

Summary

The application of PBIS to preschool settings is relatively new and has increased in implementation with support from centers as CSEFEL and Technical Assistance Center on Social Emotional Intervention for Young Children (TACSEI). There are several differences in the application of PBIS between school-aged and early childhood settings. In order to support a developmentally appropriate model of PBIS for young children, several adaptations have been made to the terminology, behavioral expectations, definitions and responses to challenging behaviors, and data collection systems that are used in early childhood settings (Fox & Little, 2001; Muscott, Pomerleau, & Szczesiul, 2009). Several gaps in the literature focusing on young children with challenging behaviors in typical settings remain. To date, there is limited research assessing the effectiveness of a tiered program-wide intervention in settings serving young children. Further, few studies have investigated student outcome data with comparison to baseline or a control group. In addition, more research is needed to understand how teacher implementation fidelity may affect student outcomes. The purpose of this investigation was to examine the effectiveness of PWPBIS across multiple early childhood sites. An additional purpose was to examine the role of implementation fidelity on child outcomes.

Chapter Three

Methodology

Participating Programs and Children

The data for this study were collected as part of a state Early Intervention Positive Behavior and Intervention Support (EIPBIS) grant funded by the Bureau of Early Intervention Services and Office of Child Development and Early Learning. Competitive funding allotments were distributed to selected state public education agencies that submitted a complete grant application. The state public education agencies that received grant funding made a commitment recruit preschool programs whose staff would agree to attend PBIS training sessions, implement PBIS procedures, collect data, and report data to the state. Twelve state public education agencies that support programs servicing young children received grant funding in 2010-2011. Within the 12 state public education agencies, 59 programs were selected to participate and received grant support. In total, 63 programs collected social emotional data. Of those, 59 programs received grant support and four did not receive grant support, 52 were excluded from this study because they either used alternative assessments, used inconsistent scales (one rating scale at pre-implementation assessment and a different scale at post-implementation assessment), or decided not to participate. Of the remaining seven programs, one program dropped out and six programs were included in this study. A detailed description of the participating programs is provided in Figure 1. The state agency responsible for the grant implementation permitted the researcher access to data; however, individual child demographic data were not provided.

Data were obtained from six of the seven programs that evaluated student outcomes using the Devereux Early Childhood Assessment (DECA; LeBuffe & Naglieri, 1999) and teacher fidelity using The Teaching Pyramid Observation Tool for Preschool Classrooms (TPOT;

Hemmeter, Fox, & Snyder, 2008). The remainder of the programs used the Preschool and Kindergarten Behavior Scales - Second Edition (PKBS; Merrell, 2003) and The Social Skills Improvement System (SSIS; Gresham & Elliot, 2008); however, there were no comparison schools that used these measures. Thus, a limited number of preschools were included to assure consistency of outcomes and measurement of implementation for comparison and experimental purposes. Seven programs used both the DECA and TPOT. Data from one program were not included because the internal coach left during the implementation year and the program discontinued involvement. In addition, data from all of four programs were collected that administered the aforementioned assessments but did not implement PWPBIS.

Data from a total 10 preschools from four public state education agencies located in the Northeast were used for the current analysis. Of the participating programs, four were community private preschools (all in the intervention group) and six were Head Start preschools (two in the intervention group and four in the control group). A total of 58 classrooms participated (34 intervention and 24 control). Teachers in all of the classrooms held certification in Early Childhood Education or a related field. A total of 743 children (349 intervention and 394 control) were included as participants in this study.

The data set provided was for subjects with complete data. Demographic information by program is provided in Table 1. In addition, a z test was calculated to compare the mean percentages between the intervention group and the control group on the three demographic variables: (a) minority, (b) IFSP/IEP (percentage of children receiving Individualized Family Support Plan/Individualized Education Plan services), and (c) gender. Programs self-selected to be included in intervention or to not receive intervention. A total of 21 classroom teachers received training in PBIS. The control group was identified by the state grant facilitator as a

convenience sample of 24 classrooms from four Head Start programs in one regional area that were assessed using the DECA but did not receive training and did not implement PBIS.

Measures

Student Outcome. The Devereux Early Childhood Assessment (DECA; LeBuffe & Naglieri, 1999), a strength-oriented assessment of the social emotional health of preschoolers ages 2 to 5, was administered in each preschool program in an effort to gain student information regarding social emotional needs. The DECA, is a 37-item norm referenced rating scale based on a protective/risk factors framework. In this study, the teacher rating form was completed for all students in each preschool program. On the rating scale, 27 of the items measure one of three protective factors, Initiative, Self-control, and Attachment, which sum to provide a Total Protective Factor (TPF) score. On the protective factors subscales and TPF, higher scores indicate greater social emotional skills. The highest score that can be obtained on the TPF is 101 and above. Ratings of 55 and below on the TPF indicate the need for more intensive supports for the child. The remaining items measure a factor labeled Total Behavior Concern (TBC). On the TBC, lower scores indicate lower rates of behavioral concern. Scores on the TBC can range 26 and above to one. Raw scores at 15 or above indicate that the child may need additional supports. For the purpose of this study, the TPF and TBC raw scores were used. The pre-test DECA was completed by the teachers for students attending the preschool in May 2010 and for new students to the preschool in December 2010. The post-test DECA was completed by the teachers in May 2011.

The DECA is an adequately reliable and valid instrument. Reliability statistics for the protective/risk factors are high with the following internal reliability coefficients alphas reported: Initiative, .90, Self-control, .90, and Attachment, .85. Test-retest reliability and inter-rater

reliability for teacher-teacher pair coefficients r were reported for each respectively: Initiative, .91 and .59, Self-control, .91 and .77, and Attachment, .87 and .57. The Total Protective Factors reliability scores are high with an internal reliability coefficient alpha of .94, The Behavior Concerns sub-scale reliability scores are low with an internal reliability coefficient alpha of .80, test-retest reliability r of .68, and inter-rater reliability coefficient r of .62. The coefficient alphas for construct validity and criterion validity are reported to be .65 and .69, respectively (LeBuffe & Naglieri, 1999).

Teacher implementation fidelity. *The Teaching Pyramid Observation Tool for Preschool Classrooms* (TPOT; Hemmeter, Fox, & Snyder, 2008) is a rating scale that measures implementation of positive behavior support and classroom quality in early childhood settings. Use of the TPOT requires direct observation of EIPIBS implementation for a minimum of 2 hours including at least one child-directed routine (e.g., free play) and one teacher-directed routine (e.g., circle time). In addition, a structured interview session is conducted with the classroom teacher related to implementation of the Teaching Pyramid model. The tool is comprised of 38 items with questions that vary in response format, including a dichotomous yes/no response based on observation (items 1-7), rating scales based on teacher observation and interview (items 8-18) ratings based primarily on teacher responses to interview questions (items 19-22), and a checklist record of any “red flags” or issues that may impede performance in the classroom (items 23-38) (Hemmeter et al.). The TPOT represents the three tiers of the Teaching Pyramid model: universal, secondary, and tertiary strategies (Hemmeter & Fox, 2009). The indicators measured in the TPOT represent environmental features, ratings of instructional practices, and a list of red flag items that bring forth teaching practices or classroom issues that may be detrimental to the promotion of effective practices. The TPOT was completed by the

internal coaches for each teacher providing intervention in a classroom. A total score was obtained by adding the results of items 1-22. The highest possible score that can be obtained on the TPOT is 82. The TPOT scores used in this study were taken in May of 2011, at the end of the intervention year. The psychometric properties of the TPOT are currently under investigation and have not been published to date.

Procedures

Personnel training and support. In preparation and throughout EIPBIS implementation, each program utilized multiple levels of training and support for personnel. A coaching structure was implemented to disseminate information and support teachers. Representatives of a state public education agency, the Intermediate Unit, appointed an external coach to service multiple programs in a specific area of the state. All of the external coaches had graduate degrees, prior experience training staff on behavioral interventions, and had been involved in adapting the pyramid model for early childhood settings. Each external coach had the primary responsibility of management of the grant implementation for the sites, meeting with the core team members and internal coaches one time per month, and helping coaches to implement and facilitate trainings for teachers.

The preschool program directors each identified an internal coach to support implementation of PBIS. The internal coaches had bachelors and graduate degrees and experience implementing behavioral interventions in the classroom. All but one of the internal coaches worked as a teacher or consultant in the early childhood program that it was serving as the internal coach. The internal coach had the primary responsibility of disseminating information to the preschool teachers of one program. The internal coach planned and facilitated trainings for the program team and coached and provided feedback to program teams and

teachers. The internal coaches were trained in a number of ways. In 2009, prior to implementation, the internal and external coaches attended a two-day EIPBIS early intervention strand. In addition, the internal and external coaches attended a four-day Professional Development Instructors Institute (PDII) in July of 2010. The coaches received a strong foundation in the pyramid model with presentations of research from national Early Intervention Technical Assistance Consultant and Early Childhood Mental Health Consultant experts. Then in early August of 2010, external and internal coaches attended a 2-day training on Coaching for Effective Program-Wide Implementation of Inclusion as it applies to implementation of the Pyramid Model. This training was exclusively offered to coaches facilitating the implementation of the pyramid model and was presented an individual with extensive experience in the field. Thereafter, leaders at the state level arranged conference call meetings with coaches that occurred twice a month. In addition, the external coaches held monthly meetings with internal coaches.

Teacher training was provided by external and internal coaches. CSEFEL training Module 1, Promoting Children's Success: Building Relationships and Promoting Supportive Environments, was presented to teachers the last week in November 2010. Preschools in Program C (refer to Table 1) were trained earlier, March 2010. Module 1 covered CSEFEL universal support strategies, including topics such as designing the physical environments, building relationships, schedules/routines/ transitions, giving directions, and classroom rules. CSEFEL Module 2, Promoting Social Emotional Competence, was presented to teachers including targeted interventions such as identifying the need to teach social skills, developing friendships, enhancing emotional literacy, and problem solving. Module 2 was presented to teachers in February 2011. Preschools in Program C were trained on Module 2 earlier, in

November 2010. At the program level, internal coaches provided monthly trainings. Monthly training included topics such as enhancing the environment using visuals to increase predictability and create structure within the classroom, collecting information on the DECA and Behavior Incidence Reports (BIRS), establishing and reinforcing classroom rules using an expectations matrix, and embedding strategies into lesson planning,. In addition, internal coaches joined teachers in the classroom and met with them weekly for coaching and feedback. The number of hours of coaching and feedback varied from teacher to teacher, averaging one hour per week per teacher. The researcher was unable to gain access to individual teacher coaching and feedback times. Training on the DECA was provided by the external coach, who had been previously trained by the Devereux Organization. The DECA training for the internal coaches was approximately an hour in length. Internal coaches participated in an additional DECA scoring training that lasted approximately two hours.

Data Collection. The DECA rating scales were completed pre- and post-intervention by each child's teacher. Administration took place prior to implementation of EIPBIS intervention strategies in May 2010 and again at the end of the year of implementation, May 2011. In some cases, when the student did not attend the program the year prior, the pre-intervention DECA was administered after entry into the program (November and December, 2010) which coincided with the initial Module 1 training.

Teacher fidelity data for this study were collected by the internal coaches at the end of the intervention year. The internal coach was trained in the TPOT tool by participating in a webinar developed by two of the authors of the TPOT, Drs. Hemmeter and Fox, in conjunction with the Technical Assistance Center on Social Emotional Intervention for Young Children (TACSEI) and Center on the Social Emotional Foundations of Early Learning (CSEFEL). The webinar

reviewed every section of the tool in detail with video examples of teacher/student interactions to illustrate and to practice scoring. The webinar was approximately an hour in length. On initial TPOTs, the external and internal coaches both completed the assessment tool, compared responses and came to consensus. Formal interobserver agreement data were not reported and the method of obtaining consensus between observers was not clear.

Teacher years of experience and education level data were obtained from the external coaches and internal coaches. The external coaches and internal coaches reported the years of experience and education level of each of the teachers that implemented during the 2010-2011 school year.

Experimental Design, Data Analysis, and Interobserver Agreement

A quasi-experimental pre/post research design was used to evaluate student outcomes. Data were entered into Excel and entry accuracy was checked for 30% of randomly selected sessions by having a second individual compare the entered score with the assessment score. The Excel document was then transferred to SPSS for analysis.

The following statistical analyses were conducted to answer each of the research questions. For the first research question, examining the growth between students that participated in PWPBIS versus students that did not participate in PWPBIS, a mixed analysis of variance (ANOVA) was conducted to simultaneously examine the growth from pre- to post- (time) and between treatment and control (groups). In the presence of a statistically significant interaction, simple effects were conducted, using an independent samples t-test, to compare the condition effects at each time point. Data for Question 1 included data from intervention programs 1-4 and control programs 1-4 (Table 1) with a total sample size of 703 children. Data from intervention programs 5 and 6 were not used because training occurred prior to the

collection of baseline. A power analysis, using GPower (Faul & Erdfelder, 1992), indicated that a sample size of 200 would be sufficient to detect a statistically significant interaction effect for the time-by group interaction with power of .80 assuming $\alpha=.05$ and a small effect size ($f=.10$). For the within subjects, time (pre/post) parameter, a minimum sample size of 200 was required to achieve a power of .80 assuming $\alpha=.05$ and a small effect size ($f=.10$). For the between subjects, condition (intervention/control) variable a minimum sample size of 98 was required assuming $\alpha=.05$ and a medium effect size ($f=.25$) to achieve a power of .80.

For the second research question, examining whether teacher variables (fidelity of implementation, years of experience, education level) predicted post-test child outcomes while controlling for pre-test child outcomes, hierarchical linear modeling (HLM) was used for the statistical analysis. Data from all programs that received training were included in the analysis for Research Question 2, with a sample size of 349 students and 21 teachers. Teacher fidelity was the primary variable of interest. Teacher years of experience and education level were added as covariates to account for their contribution in the variability in student outcomes that could not be explained by the variable of interest.

Nested or hierarchical data structures are used when information is gathered at the individual level and at another level such as the classroom or school level. When nested structures exist, it is often the case that the groups are more similar to each other than if they were randomly sampled (Osbourne, 2000). For example, students in one classroom may be more similar to each other than students sampled from a larger population. In that classroom, experiences are being shared and being in the same environment over time may lead to increases in homogeneity and, therefore, these individuals are not fully independent. Although, HLM is similar to Ordinary Least Squares regression (OLS) an outcome variable is predicted as a

combination of one or more variables and the intercept. Therefore, it was an appropriate analysis for this study because it accounts for the effects of level 2 variables on the outcome and can model the cross-level interaction, recognizing the partial independence of individuals within the same group (Hofmann, 1997; Osbourne, 2000).

The HLM analyses involved two levels. Level 1 was the individual student level and level 2 was the teacher level. Of specific interest was the relationship between student outcome scores (level 1) and fidelity of implementation (level-2 predictor). In addition, the predictors of teacher years of experience and teacher's education level were explored as level 2 variables. Model testing proceeded in three phases: unconditional means model to calculate the intraclass correlation coefficient (ICC), random intercepts and random slopes model, and random intercepts only model. The unconditional means model calculates the variance in student outcomes attributable to teacher membership. Next, the random intercept and random slope model or full model was conducted to minimize the likelihood that a type 1 error may occur because individuals are not fully independent. A reduced model, random intercepts only model, was the final model expressed.

Interobserver Agreement

Interobserver agreement was calculated for both entering of the data from the actual rating scales to the excel file or transferring electronic information to an excel file and for scoring of the rating scales. The scores rated for interobserver agreement of scoring and entering were randomly selected. First, data were entered by the researcher and 30% of the scores entered for each classroom for intervention programs 1-4 and control programs 1-4 were checked by a second individual who was a graduate student in Special Education. Next, 30% of the rating scales in each classroom that were available (intervention programs 1 & 2) were scored again to

check for accuracy. Rating scales were not rescored for intervention programs 3 and 4 and control groups 1-4 because the actual rating scales were not available to the researcher (scores were sent electronically). For both levels, interobserver agreement was calculated on by checking the total scores for both the TPF and TBC subtests of the DECA rating scale, dividing the number of agreements by the total number agreements plus disagreements and multiplying by 100. Inter-observer agreement on entering data for program 1 was 94.24% (range, 92-100%), program 2 was 98.8% (range, 97.5-100%), program 3 and 4 were 100%, and programs 7-10 were 100%. Interobserver agreement on the scoring accuracy for program 1 was 98.8% (range, 96.9-100%) and program 2 was 97.9% (range, 95.9-100%).

Chapter 4

Results

Table 1 reports program and school characteristics. Z scores were computed for the variables of minority status, gender, and having an IFSP. A significant result of a two-tailed z test is indicated with a p value of .025 or less. Results of z test for minority status indicated a non-significant z score of 6.5 ($p = .51$), gender had a non-significant z score of .46 ($p = .46$), and IFSP had a non-significant z score of .04 ($p = .14$), indicating there were no group differences on these variables.

RQ 1: Do post-test outcomes of students from preschools that participated in PWPBIS across an academic year show significantly greater growth from pre-test to post-test on a standardized measure of social emotional skills than children in preschool classrooms that did not participate in PWPBIS?

Descriptive statistics were calculated and are summarized in Tables 2. A mixed ANOVA was used to examine the differences in growth between time (pre and post), condition (intervention and control) and the interaction between time and condition, on two social emotional subscales of the DECA (TPF and TBC). Prior to conducting the analysis, data were examined with regards to meeting statistical assumptions of the mixed ANOVA procedure. First, univariate normality was within the acceptable ranges of +2 and -2 based on skewness for the pre-TPF (-.559) and skewness for the post-TPF (-.494) and kurtosis (.361) for the pre-TPF and kurtosis (.432) for the post-TPF. The skewness of the pre-TBC (5.146) and post-TBC (2.553) and the kurtosis for pre-TBC (46.628) and post-TBC (16.027) had positive values above the acceptable range, indicating violations. Second, the Levene's test of equality of error variances revealed that the assumption of homogeneous covariance matrices of the dependent

measures of each group on the TPF were violated ($p < .001$) and on the TBC were violated ($p < .001$) indicating the variances were significantly different.

The data were transformed using a log transformation to account for the violations in assumptions. The statistical assumptions of the transformed data from both the original and the transformed data are reported in Table 2. As a result of the transformation, the TBC subscale was within the acceptable ranges for skewness, kurtosis, and homogeneity of variance. The TPF subscale violated the kurtosis and homogeneity of variance assumptions. Although there were improvements in the assumptions for the TBC subscale, after conducting the mixed ANOVA with the transformed data, similar outcomes were found for both the TPF and TBC. On the TPF subscale with the transformed data, there was a statistically significant difference between scores at the two times (pre and post), regardless of condition $F(1, 756) = 82.551, p < .001$; and there was statistically significant main effect for condition, $F(1, 756) = 7.160, p = .009$, indicating that ignoring time, there was no difference between the ratings in the two conditions (intervention versus control). In addition, the interaction between time and condition was significant, $F(1,756) = 31.500, p < .001$.

On the TBC subscale with the transformed data, there was a significant main effect for time was found, $F(1,756) = 37.291, p < .001$, indicating that there was a statistically significant improvement between scores at pre and post. A significant main effect for condition was found, $F(1,756) = 15.113, p < .001$, indicating that ignoring other variables, there was a statistically significant difference between the ratings in the two conditions (intervention versus control). A non-significant effect was found for the interaction between time and condition, $F(1,756) = 1.635, p = .201$. The outcomes of the original analysis and the transformed analysis are

consistent. The original outcomes were used for final analysis because the original data are more meaningful for interpretation.

Using a 2 time (pre and post) X 2 condition (intervention versus control) mixed ANOVA design for Total Protective Factor (TPF) subscale as implemented with SPSS GLM, there was a significant main effect for time, $F(1,701) = 84.04, p < .05$, indicating that there was a statistically significant difference between scores at the two times (pre and post), regardless of condition. There was not a statistically significant main effect for condition, $F(1, 701) = 1.83, p = .176$, indicating that ignoring time, there was no difference between the ratings in the two conditions (intervention versus control). However, the interaction between time and condition was statistically significant, $F(1,701) = 46.73, p < .005$, partial $\eta^2 = .062$, which renders the significant main effect for time misleading. This significant interaction indicates that although both conditions (intervention and control) performed better on the post-test than on the pre-test, the groups varied by time, with the control group ($N = 394$) having greater gains from pre ($M=69.93$) to post ($M=76.68$) than the intervention group ($N=309$) had from pre ($M = 74.03$) to post ($M = 75.01$) (Figure 3). The two groups had significantly different improvements in outcomes scores from pre to post, $F(1, 701) = 46.73, p < .001$. The control group had a bigger growth (mean at pre = 69.93, mean at post = 76.68) than the intervention group (mean at pre = 74.03, mean at post = 75.01). Due to the significant interaction, simple effects were examined using an independent sample t -test. The time effects within condition t -test revealed a statistically significant difference between control and intervention groups at pre-test, $t(701) = 4.34, p < .001, d = .32$. The difference between control and intervention groups at post-test was non-significant, $t(701) = 1.74, p = .082, d = .13$.

Using a 2 time (pre/post) X 2 condition (intervention/control) mixed ANOVA for Total Behavior Concern Scale (TBC) as implemented with SPSS GLM, a significant main effect for time was found, $F(1,701) = 13.66, p < .005$, indicating that there was a statistically significant improvement between scores at pre and post. A significant main effect for condition was found, $F(1,701) = 39.5, p < .005$, indicating that ignoring other variables, there was a statistically significant difference between the ratings in the two conditions (intervention versus control). The interaction between time and condition was not statistically significant, $F(1,701) = .36, p < .545$, partial $\eta^2 = .001$ (Figure 4).

Question 2: Do teacher variables (fidelity of implementation, years of experience, education level) predict post-test child social emotional outcomes while controlling for pre-test social emotional status?

For the HLM analysis (Table 3), the individual student level (Level 1) had a total of 349 participants. The percentage of indicators on the treatment fidelity tool (TPOT) were calculated for each teacher and displayed in Figure 2. Fidelity of implementation scores ranged from 35 to 82. After training, a criterion level of 70-80% of implementation is expected (Hemmeter & Fox, 2009). Seventeen out of 21 teachers (81%) obtained a score of 70% or higher on the TPOT. Teacher years of experience ranged from 1 year to 24 years ($M = 11.5$). Education level ranged from having certification in early childhood education to graduate degrees. Raw scores were reported and analyzed. Data for Level 1 variables are reported in Table 4. At the teacher level (Level 2), the number of participants was 21. The level 2 predictors of treatment fidelity mean, as measured on the TPOT, years of teaching experience and education level are reported in Table 4.

Hierarchical linear modeling was used to analyze whether teacher fidelity of implementation of PWPBIS scores, years of experience, or education level (level 2 variables) predicted student outcome scores (level 1) on each of two subscales (TPF and TBC) of the DECA. Model testing proceeded in three phases: (1) unconditional means model to calculate the intraclass correlation coefficient (ICC), (2) random intercept and random slope model, and (3) random intercept only model.

First, the ICC was calculated to determine whether a large percentage of the variance in the student outcome was attributable to teacher clustering. Although the analysis was conducted at the student and teacher levels, students (level 1) were clustered within 21 teachers which could result in potential clustering effects due to classroom teacher membership. The ICC was calculated in HLM 7 (Raudenbush, Bryk, Cheong, Congdon, & DuToit, 2011) for the unconditional model with no predictors at either level 1 (student) or level 2 (teachers). The ICC was calculated with the formula, $\tau_{00}/(\sigma^2 + \tau_{00})$ (Singer, 1998). Twenty-eight percent of the variance of the TPF subscale scores and 29.3% of the variance of the TBC subscale scores were attributed to classroom teacher membership. Using HLM to control for intraclass correlation effect was justified for ICCs with higher 15% (Enders, 2009) therefore a multi-level modeling process was used to minimize the likelihood of type I errors occurring due to violating independence assumptions. The mathematic equations for the unconditional model for the TPF and TBC were as follows:

TPF Level 1 (student level): $POSTTPF_{ij} = \beta_{0j} + r_{ij}$

TPF Level-2 (teacher level): $\beta = \gamma_{00} + u_{0j}$

Combined, the TPF Mixed Model is: $POSTTPF_{ij} = \gamma_{00} + u_{0j} + r_{ij}$

TBC Level 1 (student level): $POSTTBC_{ij} = \beta_{0j} + r_{ij}$

TBC Level-2 (teacher level): $\beta = \gamma_{00} + u_{0j}$

Combined, the TBC Mixed Model is: $POSTTBC_{ij} = \gamma_{00} + u_{0j} + r_{ij}$

Next, the full model with random intercept and slope was run in HLM 7. The mathematic equations for the full model of random intercept and slope for the TPF and TBC were as follows:

TPF Level 1 (student level): $POSTTPF_{ij} = \beta_{0j} + \beta_{1j}*(PRETPF_{ij}) + r_{ij}$

TPF Level-2 (teacher level): $\beta_{0j} = \gamma_{00} + \gamma_{01}*(TF_j) + \gamma_{02}*(YRSEXP_j) + \gamma_{03}*(ED_LEVEL_j) + u_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11}*(TF_j) + \gamma_{12}*(YRSEXP_j) + \gamma_{13}*(ED_LEVEL_j) + u_{1j}$

Combined, the TPF Mixed Model: $POSTTPF_{ij} = \gamma_{00} + \gamma_{01}*TF_j + \gamma_{02}*YRSEXP_j + \gamma_{03}*ED_LEVEL_j + \gamma_{10}*PRETPF_{ij} + \gamma_{11}*TF_j *PRETPF_{ij} + \gamma_{12}*YRSEXP_j*PRETPF_{ij} + \gamma_{13}*ED_LEVEL_j *PRETPF_{ij} + u_{0j} + u_{1j}*PRETPF_{ij} + r_{ij}$

TBC Level 1 (student level): $POSTTBC_{ij} = \beta_{0j} + \beta_{1j}*(PRETBC_{ij}) + r_{ij}$

TBC Level-2 (teacher level): $\beta_{0j} = \gamma_{00} + \gamma_{01}*(TF_j) + \gamma_{02}*(YRSEXP_j) + \gamma_{03}*(ED_LEVEL_j) + u_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11}*(TF_j) + \gamma_{12}*(YRSEXP_j) + \gamma_{13}*(ED_LEVEL_j) + u_{1j}$

Combined, the TBC Mixed Model: $POSTTBC_{ij} = \gamma_{00} + \gamma_{01}*TF_j + \gamma_{02}*YRSEXP_j + \gamma_{03}*ED_LEVEL_j + \gamma_{10}*PRETBC_{ij} + \gamma_{11}*TF_j *PRETBC_{ij} + \gamma_{12}*YRSEXP_j*PRETBC_{ij} + \gamma_{13}*ED_LEVEL_j *PRETBC_{ij} + u_{0j} + u_{1j}*PRETBC_{ij} + r_{ij}$

The full model revealed that variability for the intercept of the TPF ($p=.202$) and the intercept of the TBC ($p = .053$) and slope of the TPF ($p = .205$) and the slope of the TBC ($p = .069$) of the level 2 predictor were non-significant. Therefore, a random effects model of intercept and slope was reduced to a random intercept model. The mathematic equations for the reduced model of random intercept only for the TPF and TBC were as follows:

TPF Level 1 (student level): $POSTTPF_{ij} = \beta_{0j} + \beta_{1j}*(PRETPF_{ij}) + r_{ij}$

TPF Level-2 (teacher level): $\beta_{0j} = \gamma_{00} + \gamma_{01}*(TF_j) + \gamma_{02}*(YRSEXP_j) + \gamma_{03}*(ED_LEVEL_j) + u_{0j}$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}*(TF_j) + \gamma_{12}*(YRSEXP_j) + \gamma_{13}*(ED_LEVEL_j)$$

Combined, the TPF Mixed Model: $POSTTPF_{ij} = \gamma_{00} + \gamma_{01}*TF_j + \gamma_{02}*YRSEXP_j + \gamma_{03}*ED_LEVEL_j + \gamma_{10}*PRETPF_{ij} + \gamma_{11}*TF_j *PRETPF_{ij} + \gamma_{12}*YRSEXP_j*PRETPF_{ij} + \gamma_{13}*ED_LEVEL_j *PRETPF_{ij} + u_{0j} + r_{ij}$

TBC Level 1 (student level): $POSTTBC_{ij} = \beta_{0j} + \beta_{1j}*(PRETBC_{ij}) + r_{ij}$

TBC Level-2 (teacher level): $\beta_{0j} = \gamma_{00} + \gamma_{01}*(TF_j) + \gamma_{02}*(YRSEXP_j) + \gamma_{03}*(ED_LEVEL_j) + u_{0j}$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}*(TF_j) + \gamma_{12}*(YRSEXP_j) + \gamma_{13}*(ED_LEVEL_j)$$

Combined, the TBC Mixed Model: $POSTTBC_{ij} = \gamma_{00} + \gamma_{01}*TF_j + \gamma_{02}*YRSEXP_j + \gamma_{03}*ED_LEVEL_j + \gamma_{10}*PRETBC_{ij} + \gamma_{11}*TF_j *PRETBC_{ij} + \gamma_{12}*YRSEXP_j*PRETBC_{ij} + \gamma_{13}*ED_LEVEL_j *PRETBC_{ij} + u_{0j} + r_{ij}$

The random intercept model revealed significant results of the random effects for the variability of the intercept of the TPF ($p < .001$) and the TBC ($p < .001$) of the level 2 predictor.

The random intercept model revealed similar non-significant findings for treatment fidelity predictor for the intercept of the TPF ($p = .445$) and the intercept for the TBC ($p = .447$) and slope for the TPF ($p = .424$) and the slope for the TBC ($p = .225$). The p value for intercept 1, years of experience, was statistically significant for the intercept of the TPF ($p = .008$) and for the intercept of the TBC ($p = .010$) and pretest effect ($p = .010$) and the pretest effect of the TBC ($p = .016$). Controlling for pre-test and other effects on the TPF, each additional year of teacher experience was negatively associated with a change in student outcomes. In other words, for each additional year of teacher experience, student post-test scores decreased by 1.12 units in the intercept. Holding other effects constant, an additional year of teacher experience was associated with a slight pre-test slope increase by .02 units. For the TBC, controlling for all other effects,

for each additional year of experience a teacher had, the intercept of the student post-test decreased by .186 units. The slope of pre-test was 0.258. The pretest effect is strengthened by .017 units for each additional year of experience. Teacher years of experience was associated with improved student outcomes on the TBC. The TPF and TBC intercepts and slopes for teachers' education level were non-significant, $p > .05$. The estimated HLM coefficients and other statistics are presented in Table 5 and 6.

Chapter 5

Discussion

The primary purpose of this study was to extend the literature on PWPBIS by examining the impact of training in PWPBIS on the gains on child social emotional outcomes. A secondary purpose was to examine the influence of treatment fidelity and other teacher variables (years of experience, education level) on the social emotional and behavioral outcomes of young children. Overall, results for the TPF subscale indicated that there was a significant interaction between time and condition, though it was the control group that showed more growth. The child outcomes on the TBC subscale showed that although all scores decreased, indicating improvement from pre to post-test, the interaction between time and condition was not significant. In addition, the teachers' experience variable was a statistically significant contributor to improved student outcomes on the TBC alone. Treatment fidelity and teacher's education level did not significantly contribute to student outcomes on either subscale.

PWPBIS Impact on Child Outcomes

According to pre and post-test data, child TPF outcome indicated a slight improvement in the intervention group and a large improvement in the control group. The small change for the intervention group may be explained by the difference between the intervention and control groups at pre-test. The mean pre-test score on the TPF for the intervention group started higher with less room for growth while the control group started lower and scores increased to just exceed where the intervention group had started. In addition, the mean scores for each condition were in the typical range for the TPF subscale on the DECA. The typical range on the DECA was a raw score between 56 and 85. Scores in the typical range may suggest that the sample was generally healthy and therefore a large amount of growth would not be expected.

The groups being compared were not randomized or shown to be equivalent at pre-test, and the results should be taken with caution. The difference in growth between the control and intervention groups may have been due to the variability in each child's history of exposure to social emotional skills and supports. The control group was comprised of all Head Start programs and research has shown that children from low-income families have higher rates of problem behavior (Qui & Kaiser, 2003). Therefore, it may have been the case that these children benefited greatly from being in a structured school setting that provided any form of social emotional support.

Another possibility is that PBIS does not have an impact on children's outcomes. The minimal growth between the pre- and post-test outcomes for the intervention group suggests that mean scores did not change as a result of PBIS implementation. Although, in this case, it may have been beneficial to look at individual student pre- and post-test differences rather than mean scores to determine if any effect of the intervention could be detected. By looking only at the means, more specific information about the intervention effectiveness may have been lost. For example, if individual scores were analyzed, there may have been information showing that the students with scores below a certain level showed more growth than students above that level or that the growth of students in the at-risk range moved them into a typical range on the outcome measure.

Another explanation may be that the control group teachers were implementing components of PBIS as part of their typical classroom routines. For example, in a study of school-wide PBIS implementation in 37 elementary schools Bradshaw, Reindke, Brown, and Bevans (2008) found that, at baseline, results of the SET indicated elements of PBIS were being used in classrooms in both conditions. Following implementation of PBIS, increases on the SET

were seen for both the control and intervention groups although greater increases were seen in the intervention group. In the current study, child outcomes could have been affected if elements of PBIS were used by the teachers in the control group. For example, there may have been universal strategies in place in the control condition. Universal supports are put into place to support young children to learn to solve problems, understand emotions, and build relationships (Joseph & Strain, 2003; Smith, Lewis, & Stormont, 2011) and it would make sense that these skills would be reinforced regularly in early childhood settings. It would have been beneficial in this study to have had information on the implementation of PBIS strategies in each classroom at pre-test for all of the classrooms.

On the TBC subscale, students in the intervention group did not make greater gains from pre- to post-implementation than students in the control group. Both condition group means improved from pre to post. Therefore, the results do not indicate whether the intervention contributed to the improvement in scores. This could be due to several factors.

The quality of intervention delivery may have affected the behavioral outcomes. In order for children to learn rules, expectations, and the skills to be successful with other children, they need to be taught explicitly. Behavioral rules and expectations need to be reinforced often and consistently. Although the TPOT may capture some of what is happening in the classroom, it is unclear whether the children are getting the dosage of behavioral support that is necessary to make a change in the behavioral concern score. For example, on the TPOT it may be recorded that behavioral expectations are posted and reviewed during large group activities. But there may be some children that need the expectations stated more frequently, especially if the expectations are new to them or if it is prior to a situation that may trigger a behavior. These subtle differences in the quality of implementation are not addressed.

Another possibility is that the TBC scores did not decrease substantially because more intensive or individualized strategies were necessary to address the student's behavior. For example, if students were being rated as having poor impulse control and that behavior was not reducing with universal supports or targeted supports, the student may have needed tertiary (individual) supports. Training for tertiary supports was not included in the 2010-2011 year. The external coaches reported time as a barrier to implementing all three tiers. Training and implementation of universal and targeted supports took more time than was anticipated.

In addition, it may be the case that when looking at behavioral concern, the assessment needed to be given at a midyear point as well so that teachers could adjust or plan more intensive interventions. A component of PBIS is conducting on-going monitoring of student's progress (Sugai et al., 2000). By only having assessment at the beginning and end of the year, the teachers may have been missed opportunities to address student's needs.

The significant condition main effect indicated that ignoring time, there was a difference between the scores of the control group and the scores of the intervention group. One explanation for the difference in the groups is that the programs were not randomly assigned to either intervention or control condition. This lack of randomization is a threat to the internal validity of the study. In addition, although both groups made gains on the TBC, the control group had consistently higher behavioral concerns at pre-test and post-test than the intervention group. The difference in groups was apparent even though growth occurred. This group difference was seen on the TPF subscale as well. A significant simple effect was found for the TPF at pre-test. Overall, the lack of equivalency between the intervention and control groups makes it difficult to attribute the improvement in scores to the intervention.

Teacher Variables

The primary teacher variable of treatment fidelity was not significantly associated with students' post- implementation social emotional outcomes while controlling for pre- implementation scores. Although much of the research supports that higher fidelity of evidence-based intervention is associated with better student outcomes (Noell, Gresham, & Gansle, 2002; Vollmer, Roane, Ringdahl, & Marcus, 1999; Wilder, Atwell, & Wine, 2006) there are a number of studies that have shown lower fidelity was associated with behavioral change as well (Arkoosh et al., 2007; Northup et al., 1997).

One possible explanation for the non-significant results may be due to implementation procedures. It is unclear whether the training on administration of the TPOT was adequate. A training criterion was not established to demonstrate adequate mastery of the assessment tool. Although the external coaches all reported that two observers implemented the assessment on several of the fidelity measures, there was no formal documentation of how many assessments were conducted by two observers and interrater agreement was not calculated.

The ratings on the TPOT were done by the internal coaches, who also trained and supported the teachers, so there may have been elevated TPOT scores due to experimenter bias. Thus, true variability in implementation may not have been measured. The TPOT scores lacked variability and the restriction in range might have affected the results. In addition, there was no record of when the TPOT administration training occurred in relation to the post-implementation TPOT ratings. Therefore, there is limited information to ensure the observers were consistent in their TPOT ratings or whether there was observer drift due to the time gap from training to administration. In addition, the TPOT does not have published psychometric properties to date. It is unclear whether the results would be similar under consistent conditions, or whether the scale is measuring what it is intended to measure. Accurate treatment fidelity information is

needed to discern whether findings are due to implementation inaccuracy or intervention ineffectiveness (Gresham, 2009).

The relationship between treatment fidelity and child outcomes is not straightforward and is influenced by the multiple dimensions of treatment fidelity (content, quality, quantity, and process) (Hagermoser Sanetti & Kratochwill, 2009). The quantity of treatment, or how much of the intervention was provided, is not captured using the TPOT and may be important to students' acquisition of targeted skills. In addition, the specific content and quality of implementation of the strategies are not necessarily captured. Implementation quality may have varied considerably across classrooms and programs. Another possibility is that the treatment fidelity score does not accurately reflect critical elements of PBIS. Therefore, a lower treatment fidelity score may not be associated with a lower outcome score. More specific information on the content being delivered may be important, as is additional information on the psychometrics of the TPOT.

In this study, 17 of 21 teachers implemented with fidelity at least 70% of the time ($M = 82\%$). It is unknown whether teacher fidelity improved with training because pre-implementation TPOT scores were not available. Hemmeter and Fox (2008) reported that teachers who are not trained in PBIS demonstrated approximately 40% of the total possible indicators on the TPOT (Hemmeter & Fox, 2008) and, therefore, the teachers in this study may have increased their percentage of implementation in the first year. In addition, other studies have shown increased levels teacher fidelity over several years of implementation (Bradshaw, Reindke, Brown, Bevans, 2008; Muscott, Pomerleau, & Szczesiul, 2009) suggesting that it would be beneficial to analyze data for subsequent years of implementation.

Although fidelity of implementation was not associated with outcomes, years of teaching experience demonstrated a statistically significant association with student outcomes.

Specifically, on the Total Protective Factors Scale (TPF), an increased score indicates an improved outcome but in this case the additional year of a teachers' experience was associated with a decreased rating in student performance (by 1.12 units). In other words, when everything else is held constant and students are similar at pre-test, if a teacher has even a year more of experience than another teacher, then the more experienced teacher would see decreased average post-test scores (1.12 units for TPF). A possible explanation for these results may be that more experienced teachers are better able to evaluate students' social abilities, which require active problem solving and social responsiveness in novel situations. Experienced teachers may credit students with smaller increments of growth, understanding that it may be difficult for students to internalize and demonstrate autonomy in the use of these skills. Thus, more experienced teachers are likely to assign lower ratings to students on this measure.

Another possibility is that teachers with more years of experience are using strategies that were learned early in their careers that are not consistent with the positive behavior supports framework. One example is that they may be implementing punitive strategies that result in suppression of behavior problems but do not teach adaptive skills. If the more experienced teachers had different philosophies about working with children with challenging behaviors, it may be more difficult for them to adopt positive strategies.

There was a difference between the outcomes of the two scales measuring behavior and protective factors. On the Total Behavior Concerns Subscale (TBC), a lower score indicates an improved outcome. For this subscale, an additional year of teaching experience was associated with improved behavioral rating (by .19 units). It is possible that teachers with more years of experience have a greater tolerance for students who exhibit social emotional difficulties or, conversely, are better at managing student behavior. Data from the current study may suggest

the latter. Specifically, the negative correlation of an additional year of experience with lower scores on the TBC may indicate that experienced teachers are applying successful behavior management strategies more often or more effectively than teachers with less teaching experience

Another explanation for the findings is that it may be that it is harder to make gains on one subscale than on the other subscale. Specifically, it may be more difficult to demonstrate improvement on protective factors than on behavior concerns. If so, it may be important to identify measures that are sensitive to change. Yet another explanation is that the difference in scores may be due to instrumentation. There are differences in the subscales, such as fewer questions on the TBC than on the TPF. In addition, it is important to consider that the slight decreases, approximately 1 unit on the rating scale for each additional year of teaching experience may not be meaningful. The small effects in this study were consistent with the research of Pianta et al. (2005) and Early et al. (2007) and help to confirm that teacher variables (including education, experience and fidelity) are complex and multidimensional. Additional research is needed in the area of treatment fidelity regarding the content, quantity, quality, and processes used in classrooms so that practitioners can better link their classroom practices with the student outcomes. In addition, investigations are needed to further understand the amount of additional training or specialization and the perspectives of the teacher' that best guide instruction.

Limitations

Several limitations need to be considered while interpreting the results of this study. First, it is important to recognize that the PWPBIS training and support efforts reported in this study were coordinated and facilitated by the state team, not the researcher. The researcher was

only involved with the aggregation and coordination of the data for analysis. Although these findings suggest that the state's efforts did not produce a greater improvement in outcome scores from pre- to post-implementation on either measure, it is unclear what teacher practices were responsible for the change that did occur. If in fact, change was due to implementation of PBIS components in either the treatment or control classrooms, then it would be interesting to examine what level of training and support may be necessary to bring about the most significant changes. This study reports the descriptive data of PBIS implementation with information aggregated through interviews and review of notes kept by program personnel. The external coaches verified that each program had attendees at the initial state training sessions and the monthly call meetings held by the state. However, detailed information on the types and amount of support provided to the coaches and teachers was not available. In addition, individual student demographic information would have been valuable but was not available to the researcher. It would have been beneficial to investigate child and teacher gender, age, and ethnicity as covariates of the outcome scores.

Second, the programs in this study were not randomly assigned to intervention and control conditions. In addition, because the programs voluntarily participated and had knowledge that PWPBIS would be implemented, the results may have been biased. It is possible that there were differences in groups because the groups became involved in the study in two separate ways. The intervention groups applied and obtained grant funding to support PWPBIS implementation. The control group did not apply for grant funding and therefore may have differed for reasons that led to their decision not to apply for grant funding.. That is, the groups may have been heterogeneous at the start of the study. Statistically significant differences were found between the intervention and control groups at pre-test and violations were found in the

assumptions for the ANOVA analysis. This threatens the internal validity of the study and may explain the difference in student outcomes between pre- and post-implementation for each group. Ideally, the groups would have been similar across all of the demographic variables at pre-test.

Third, the volunteer status of the intervention classrooms may not have captured a representative sample of children and teachers. In addition, the control group voluntarily collected the assessment information and was a sample of convenience. All of the control classrooms were Head Start classrooms from one regional area, thus limiting the results because a large heterogeneous sample of young children was not represented. Thus, the external validity was compromised and therefore limits the generalizability of the findings.

Fourth, regarding the assessment tools used, the DECA rating scale was the sole measure for assessment of social emotional outcomes. Using multiple methods of measurement would have contributed more information about outcomes and would give a broader picture of student outcomes. In school-wide PBIS, office disciplinary referral (ODR) data are often used to assess outcomes. In PWPBIS, behavioral incidence reports (BIRs) are often used by teachers to record specific behavioral incidences. Though, there is no universal procedure for recording or aggregating BIR data.

Fifth, raters of both measurement tools (DECA and TPOT) may have been bias, potentially introducing a threat to the internal validity of the study. Teachers in the intervention group, with knowledge of the purposes of the intervention, rated the children in their own classrooms on the DECA potentially introducing bias. On the TPOT measure, the internal coaches, who were training and coaching the teachers, were then rating the teachers on fidelity of implementation. External coaches reported having two scorers rate the teachers using the TPOT during the training phase but no formal data were taken.

In addition, the TPOT fidelity tool assesses all three levels of tiered implementation. Even though only three of the items on the TPOT address tertiary supports, it is still important to note that the programs in this study were only trained on universal and targeted supports. In addition, the psychometric properties of the TPOT are unknown. It would have been beneficial to use a tool that has published information regarding reliability and validity. An example would be the Pre-SET (Steed & Webb, 2012).

Implications for Research and Practice

There are several areas of research suggested as result of these findings. First, although the results of this study should be taken with caution based on the limitations, the pre-post score differences seen for each group show that further research in PBIS for young children is warranted. It would be interesting compare strategies used in control and intervention classrooms and investigate if there are specific social emotional or behavioral strategies that are associated with the greatest child gains. It is important to continue to investigate the differences and similarities in practices and outcomes of schools implementing PBIS and those schools who are not implementing PBIS. This is particularly important given the growing emphasis on PBS in elementary schools and the rapidly expanding implementation in preschools.

Second, it is important to consider efficiency and effectiveness when planning interventions. Research suggests that even with high levels of training, many teachers do not implement interventions with accuracy (DiGennaro, Martens, & McIntyre, 2005) and that as the complexity and amount of demands increase, the level of treatment integrity decreases (Lane, Bocian, MacMillian, & Gresham, 2004). Even though, the findings of this study did not support previous research showing that stronger fidelity of implementation produced the most improved student outcomes, there are many unanswered questions regarding the influence of treatment

fidelity on intervention outcomes. There continues to be a need to address the influence of treatment complexity on fidelity of implementation, especially for tiered supports. On a practical level, teachers need to understand how to use fidelity information in order to make data-based decisions regarding the most effective intervention components (Sanetti, Dobey, & Gritter, 2012). Further, in the case of complex yet effective interventions, additional research on training and coaching procedures to increase fidelity is needed.

Third, some of the limitations that threatened the internal validity of the current study were related to feasibility and ease of implementation of PBIS and assessment for the programs. For example, although best practice dictates assessing treatment fidelity using a psychometrically sound instrument, including direct observation, and controlling for experimenter bias, these procedures were not followed in the current project for reasons of feasibility. For example, in this study, pre-test TPOTs were not completed for all of the programs and documentation of training and support for the use of the TPOT tool was not available. It may be more feasible for early childhood programs to use a self-monitoring and self-reporting strategy to assess treatment fidelity, although research suggests that self-report methods may inflate the estimated levels of integrity as compared to direct observation (Lane et al., 2004). Gaining a better understanding of the feasibility of programming and assessment in early childhood settings would allow for a reconceptualization of the program wide supports that strives for an efficient, effective, feasible, and sustainable process.

There are ways in which this research can inform practice as well. First, early childhood programs can use what is already working as a starting point for implementation of effective practices. The TBC scores at baseline fell outside of the risk range and improvements for both the intervention and control groups on the TBC may suggest that the programs were using

effective strategies. Programs should identify what they are doing, what is working, and then proceed to shape their practices. Such an approach would be most efficient. Second, in order to facilitate the PBIS framework in early childhood programs, it would be beneficial to identify each program's resources and create systems within the program itself to ensure that effective assessment and intervention occurs.

Third, an item analysis of the DECA may provide information to help individual students who need it the most, resulting in greater gains. Further, although normative data can be useful, it may be beneficial for programs to use a combination of assessment tools. An assessment tool used to guide intervention is most useful when it is sensitive to change and may help to give information about skill achievement within a classroom (Miller, Linn, & Gronlund, 2013).

Therefore, programs may also add a behavioral assessment to work in conjunction with a norm-referenced assessment. For example, direct observation may be more helpful for analysis, and enable ongoing monitoring to inform programming. This is supported by data in the current study showing that the TPF did not result in change which may have been a result of instrumentation. Previous studies implementing PWPBIS all included some form of direct behavioral observation or recording of behavioral incident reports. As PWPBIS becomes more established, consistent methods of direct observation may be helpful in order to better compare results and outcomes across studies.

In many ways, this investigation pointed out that there is a gap between research and practice. The limitations that were brought forth seemed to be as a result of lack of collaboration by the state and researchers that would allow us to consider what would be necessary for implementation in order to use the data to inform future practice. The results of this study suggest, a promising approach may be that researchers and practitioners could connect and

continue to work together to establish feasible and acceptable methods to measure intervention implementation and child behavior change, as well as demographic variables that affect child outcomes.

Conclusions

There is a continued need for research in the area of PWPBIS. There are a limited number of published investigations of tiered supports for young children or the implementation of PWPBIS. This study added to the literature on PBIS by investigating the differences between control and intervention programs on student social emotional outcomes. In addition, this study, added to the literature investigating the association between treatment fidelity and student outcomes. Given that both control and PBIS implementation groups made gains pre- to post-implementation on a social emotional student outcome measure, further research exploring the effectiveness of PWPBIS is needed. In addition, in light of outcomes indicating that teacher fidelity was not associated with improved student outcomes, but teacher years of experience teaching was, the specific variables that contribute to improved student outcomes warrants additional study.

References

- Anderson, C., Lewis-Palmer, T., Todd, A., Horner, R., Sugai, G., & Sampson, N. (2008). *Individual student systems evaluation tool*. OSEP TA-Center on Positive Behavioral Interventions and Supports. Eugene: University of Oregon.
- Arkoosh, M. K., Derby, K. M., Wacker, D. P., Berg, W., McLaughlin, T. F., & Barretto, A. (2007). A descriptive evaluation of long-term treatment integrity. *Behavior Modification, 31*, 880-895. doi: 10.1177/0145445507302254
- Ashiabi, G. S. (2007). Play in the preschool classroom: Its socioemotional significance and the teacher's role in play. *Early Childhood Education Journal, 35*(2), 199-207.
- Bambara, L. M. & Kern, L. (2005). *Individualized supports for students with problem behaviors*. New York: Guildford Publications.
- Beard-Jordan, K., & Sugai, G. (2004). First Step to Success: An early intervention for elementary children at risk for antisocial behavior. *Behavioral Disorders, 29*, 396-409.
- Benedict, E. A., Horner, R. H., & Squires, J. K. (2007). Assessment and implementation of positive behavior support in preschools. *Journal of Positive Behavior Interventions, 27*, 174-192.
- Bierman, K. L. (2004). *Peer rejection: Developmental processes and intervention strategies*. New York: Guilford.
- Bradshaw, C. P., Koth, C. W., Bevans, K. B., Ialongo, N., & Leaf, P. J. (2008). The impact of school-wide positive behavioral interventions and supports (PBIS) on the organizational health of elementary schools. *School Psychology Quarterly, 23*(4), 462.

- 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.
- 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*, 133-148.
- Bradshaw, C. P., Reinke, W. M., Brown, L. D., Bevans, K. B., & Leaf, P. J. (2008). Implementation of school-wide positive behavioral interventions and supports (PBIS) in elementary schools: Observations from a randomized trial. *Education and Treatment of Children, 31*(1), 1-26.
- Brauner C. B., Stephens C. B. (2006). Estimating the prevalence of early childhood serious emotional/behavioral disorders: Challenges and recommendations. *Public Health Reports, 121*, 303–310.
- Brown, J. R., & Dunn, J. (2008). Continuities in emotion understanding from three to six years. *Child development, 67*(3), 789-802.
- Buhs, E. S., Ladd, G. W., & Herald, S. L. (2006). Peer exclusion and victimization: Processes that mediate the relation between peer group rejection and children's classroom engagement and achievement. *Journal of Educational Psychology, 98*, 1-13.
- Carr, E. G., Horner, R. H., Turnbull, A. P., Marquis, J. G., Magito-McLaughlin, D., McAtt, M. L., et al. (1999). Positive behavior support for people with developmental disabilities: A research synthesis. *American Association on Mental Retardation Monograph Series*. Washington, DC: American Association on Mental Retardation.

- Center on the Social Emotional Foundations for Early Learning [CSEFEL]. *Positive Behavioral Supports Training Modules 1,2,3a,4*. (2008). Vanderbilt University, Peabody College.
Web Site <http://www.vanderbilt.edu/csefel>
- Chitiyo, M., May, M. E., & Chitiyo, G. (2012). An assessment of the evidence-base for school-wide positive behavior support. *Education and Treatment of Children, 35*(1), 1-24.
- Cochrane, W. S., & Laux, J. M. (2008). A survey investigating school psychologists' measurement of treatment integrity in school-based interventions and their beliefs about its importance. *Psychology in the Schools, 45*(6), 499-507. doi: 10.1002/pits.20319
- Conroy, M. A., Dunlap, G., Clarke, S., & Alter, P. J. (2005). A descriptive analysis of positive behavioral intervention research with young children with challenging behavior. *Topics in Early Childhood Special Education, 25*, 157-166. doi: 10.1177/02711214050250030301
- Derby, K. M., Wacker, D. P., Berg, W., DeRaad, A., Ulrich, S., Asmus, J. et al., (1997). The long-term effects of functional communication training in home settings. *Journal of Applied Behavior Analysis, 30*, 507-531.
- DiGennaro, F. D., Martens, B. K., & Kleinmann, A. E. (2007). A comparison of performance feedback procedures on teachers' treatment implementation integrity and students' inappropriate behavior in special education classrooms. *Journal of Applied Behavior Analysis, 40*(3), 447-461.
- DiGennaro, F. D., Martens, B. K., & McIntyre, L. L. (2005). Increasing Treatment Integrity through Negative Reinforcement: Effects on Teacher and Student Behavior. *School Psychology Review, 34*, 220-231.

- Dunlap, G., Strain, P. S., Fox, L., Carta, J. J., Conroy, M., Smith, B., & Sowell, C. (2006). Prevention and intervention with young children's challenging behavior: Perspectives regarding current knowledge. *Behavioral Disorders, 32* (1), 29-45.
- Early et al., (2007). Teachers education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78*, 558-580.
- Enders, C. (2009, September). Lecture presented in Introduction to Multilevel Modeling. Arizona State University, Phoenix, Arizona.
- Fantuzzo, J. W., Bulostsky, R., McDermott, P., Mosca, S. & Lutz, M. N. (2003). A multivariate analysis of emotional and behavioral adjustment and preschool educational outcomes. *School Psychology Review, 32*, 185-203.
- Faul, F., & Erdfelder, E. (1992). GPOWER: A priori, post-hoc, and compromise power analyses for MS-DOS [Computer program]. Bonn, FRG: Bonn University, Department of Psychology.
- Flook, L., Repetti, R. L., & Ullman, J. B. (2005). Classroom social experiences as predictors of academic performance. *Developmental Psychology, 41*, 319-327.
- Fox, L., Jack, S., & Broyles, L. (2005). Program-wide positive behavior support: Supporting young children's social emotional development and addressing challenging behavior. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute.
- Fox, L., & Little, N. (2001). Starting early: Developing school-wide behavior support in a community preschool. *Journal of Positive Behavior Interventions, 3*, 251-254.

- Gilliam, W. S. (2005). Prekindergarteners left behind: Expulsion rates in state prekindergarten systems. New York, NY: The Foundation for Child Development.
- Gleason, C. (2009). An examination of changes in classroom teacher behaviors when implementing an individualized behavior support plan for a preschooler with challenging behavior. Dissertation. University of Albany.
- Greenspan, S. I. & Weider, S. (2006). *Infant and early childhood mental health: A comprehensive approach to assessment and intervention*. Washington, D. C.: American Psychiatric Publishing.
- Gresham, F. M. (2009). Evolution of the Treatment Integrity Concept: Current Status and Future Directions. *School Psychology Review, 38*, 533-540.
- Gresham, F. M., & Elliott, S. N. (1990). *Social Skills Rating System manual*. Circle Pines, MN: American Guidance Service.
- Gresham, F. M., & Elliott, S. N. (2008). *Social Skills Improvements Rating Scales*. Bloomington, MN: Pearson Assessments.
- Guess, D., Helmstetter, E., Turnbull, H. R., & Knowlton, S. (1987). Use of aversive procedures with persons who are severely disabled: An historical review and critical analysis. *TASH Monograph Series, 2*.
- Halberstadt, A. G., Denham, S. A., & Dunsmore, J. C. (2002). Affective social competence. *Social Development, 10*(1), 79-119.
- Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2009). Toward Developing a Science of Treatment Integrity: Introduction to the Special Series. *School Psychology Review, 38*(4), 445-459.

- Harden, B. J., Winslow, M. B., Kendziora, K. T., Shahinfar, A., Rubin, K. H., Fox, N. A., et al. (2000). Externalizing problems in Head Start children: An ecological exploration. *Early Education and Development, 11*, 357–385.
- Harry, B., & Klinger, J. (2006). *Why are so many minority students in special education?: Understanding race & disability in schools*. New York: Teachers College Press.
- Hawken, L. S., MacLeod, K. S., & Rawlings, L. (2007). Effects of the Behavior Education Program on office disciplinary referrals of elementary school students. *Journal of Positive Behavior Interventions, 9*, 94-101. doi:10.1177/10983007070090020601
- Hemmeter, M. L., Corso, R., & Cheatham, G. (2006). A national survey of early childhood educators: Training needs and strategies. In *Conference on Research Innovations in Early Intervention, San Diego, CA*.
- Hemmeter, M. L., Fox, L., Jack, S., & Broyles, L. (2007). A program-wide model of positive behavior support in early childhood settings. *Journal of Early Intervention, 29*, 337-355.
- Hemmeter, M. L. & Fox, L. (2009). Supporting teachers in promoting children’s social emotional competence and addressing challenging behavior. In P. J. Windton, J. A. McCollum, & C. Catlett (Eds.), *Practical approaches to early childhood professional development: Evidence, strategies, and resources* (pp. 119-142). Washington, DC: Zero to Three.
- Hemmeter, M. L., Fox, L., & Snyder, P. (2008). Teaching Pyramid Observation Tool for Preschool Classrooms (TPOT), Research Edition. Nashville: Vanderbilt University.
- Hofman, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management, 23*, 723-744. doi: 10.1177/014920639702300602

- Horner, R. H., Benedict, e. A., & Todd, A. (2005). *Preschool-wide evaluation tool*. Eugene, OR: Educational and Community Supports.
- Horner, R. H., Sugai, G., & Anderson, C. M. (2010). Examining the evidence base for school-wide positive behavior support. *Focus on Exceptional Children, 42*(8), 1-16.
- Horner, R.H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd A.W., Esperanza, J. (2009). A Randomized, Wait-List Controlled Effectiveness Trial Assessing School-Wide Positive Behavior Support in Elementary Schools. *Journal of Positive Behavior Interventions, 11*, 133-143.
- Horner, R. H., Todd, A. W., Lewis-Palmer, T., Irvin, L. K., Sugai, G., & Boland, J. B. (2004). The school-wide evaluation tool (SET): A research instrument for assessing school-wide positive behavior support. *Journal of Positive Behavior Interventions, 6*(1), 3-12.
- Irvin, L. K., Tobin, T. J., Sprague, J. R., Sugai, G., Vincent, C. G. (2004). Validity of office disciplinary referral measures as indices of school-wide behavioral status and effects of school-wide behavioral intervention. *Journal of Positive Behavior Interventions, 6*, 131-147.
- Joseph, G. E., & Strain, P. S. (2003). Comprehensive evidence-based social-emotional curricula for young children: An analysis of efficacious adoption potential. *Topics in Early Childhood Special Education, 23*(2), 65-76.
- Kaiser, A. P. (2007). Addressing challenging behavior: Systematic problems, systematic solutions. *Journal of Early Intervention, 29*(2), 114-118.
- Kendziora, K. T. (2004). Early intervention for emotional and behavioral disorders. In R. B. Rutherford, Jr., M. M. Quinn, & S. R. Mathur (Eds.) *Handbook of Research in Emotional and Behavioral Disorders* (pp. 327-351). New York: Guildford Press.

- Ladd, G. W. (1990). Having friends, keeping friends, making friends and being liked by peers in the classroom: Predictors of children's early school adjustment? *Child Development, 61*, 1081-1100.
- Lane, K. L., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2004). Treatment integrity: An essential—but often forgotten—component of school-based interventions. *Preventing School Failure, 48*(3), 36-43.
- Lane, K. L., Robertson, E. J., & Graham-Bailey, M. A. L. (2006). An examination of school-wide interventions with primary level efforts conducted in secondary schools: Methodological considerations. *Advances in Learning and Behavioral Disabilities, 19*, 157-199.
- LeBuffe, P. A. & Naglieri, J. A. (1999). Devereux early childhood assessment: Technical manual. Lewisville, NC: Kaplan Press.
- Lewis, T. J., & Sugai, G. (1999). Effective behavior support: A systems approach to proactive school-wide management. *Focus on Exceptional Children, 31*, 1-24.
- May, S., Ard, W., Todd, A., Horner, R., Glasgow, A., Sugai, G., et al. (2003). *Schoolwide information system*. Eugene: University of Oregon, Educational and Community Supports.
- Mayer, G. R. (1995). Preventing antisocial behavior in the schools. *Journal of Applied Behavior Analysis, 28*, 467-478.
- McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21*, 471-490.

- 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*, 947-959.
- McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the Journal of Applied Behavior Analysis 1991-2005. *Journal of Applied Behavior Analysis, 40*, 659-672.
- Merrell, K.W. (2003). Behavioral, social, and emotional assessment of children and adolescents (2nd ed.). Mahwah, NJ: Erlbaum.
- Miller, M. D., Linn, R. L., & Gronlund, N. E. (2013). Measurement and Assessment in Teaching (11th Edition). Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*, 247-266.
- Muscott, H. S., Mann, E., Benjamin, T. B., Gately, S., Bell, K. E., & Muscott, A. J. (2004). Positive behavioral interventions and supports in New Hampshire: Preliminary results of a statewide system for implementing schoolwide discipline practices. *Education and Treatment of Children, 27*(4), 453-475.
- Muscott, H. S., Pomerleau, T., & Dupuis, S. (2009). Anchors away! Implementing program-wide positive behavior supports at the Visiting Nurses Association child care and family resource center. *NHSA Dialog, 12*(2), 104-121.
- Muscott, H. S., Pomerleau, T., & Szczesiul, S. (2009). Large-scale implementation of program-wide positive behavioral interventions and supports in early childhood education programs in New Hampshire. *NHSA Dialog, 12*, 148-169. doi: 10.1080/15240750902774726

- Nahgahgwon, K., Umbreit, J., Liaupsin, C., & Turton, A. (2010). Function-based planning for young children at risk for emotional and behavioral disorders. *Education and Treatment of Children, 33*, 537-559.
- Noell, G. H. (2008). Research examining the relationships among consultation process, treatment integrity, and outcomes. *Handbook of research in school consultation: Empirical foundations for the field, 315-334*.
- Noell, G. H. & Gansle, K. A. (2006). Assuring the form has substance: Treatment plan implementation as the foundation of assessing response to intervention. *Assessment for Effective Intervention, 32*, 32-39.
- Noell, G. H., Gresham, F. M., & Gansle, K. A. (2002). Does treatment integrity matter? A preliminary investigation of instructional implementation and mathematics performance. *Journal of Behavioral Education, 11*, 51-67.
- Northup, J., Fisher, W. Kahang, W. K., Harrell, R., & Kurtz, P. (1997). An assessment of the necessary strength of behavioral treatments for severe behavior problems. *Journal of Developmental and Physical Disabilities, 9(1)*, 1-16.
- Osbourne, J. (2000). The advantages of hierarchical linear modeling. *Practical Assessment, Research, and Evaluation, 7(1)*. 1-3.
- Peterson, L., Homer, A., & Wonderlich, S. (1982). The integrity of independent variable in behavior analysis. *Journal of Applied Behavior Analysis, 15*, 477-492.
- Pianta et al. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science, 9*, 144-159.

- Power, T. J., Bom-Hoffman, J., Clarke, A. T., Riley-Tillman, T. C., Kelleher, C., & Manz, P. H. (2005). Reconceptualizing intervention integrity: A partnership-based framework for linking research with practice. *Psychology in the Schools, 42*, 495-507.
- Powell, D., Fixen, D., Dunlap, G., Smith, B. & Fox, L. (2007). A synthesis of knowledge relevant to pathways of service delivery for young children with or at risk of challenging behavior. *Journal of Early Intervention, 29*, 81-106.
- Quesenberry, A. C., Hemmeter, M. L., & Ostrosky, M. M. (2011). Addressing challenging behaviors in Head Start: A closer look at program policies and procedures. *Topics in Early Childhood Special Education, 30*, 209-220. doi: 10.1177/0271121410371985
- Qi, C. H. & Kaiser, A. P. (2003). Behavior problems of preschool children from low-income families: Review of the literature. *Topics in Early Childhood Special Education, 23*, 188-216.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., DuToit, M. (2011). HLM 7: Hierarchical linear modeling and nonlinear modeling. Chicago, IL: Scientific Software International.
- Raver, C. C., & Knitzer, J. (2002). Policy Paper, Promoting the Emotional Well-Being of Children and Families. "Ready to Enter: What Research Tells Policymakers About Strategies to Promote Social and Emotional School Readiness Among Three- and Four-Years-Olds. National Center for Children in Poverty: Mailman School of Public Health, Columbia University.
- Sanetti, L. M. H., Dobey, L. M. & Gritter, K. L. (2012). Treatment integrity of interventions with children in the Journal of Positive Behavior Interventions from 1999-2009. *Journal of Positive Behavior Interventions, 14*, 29-46.

- Severson, H. H., Walker, H. M., Hope-Doolittle, J., Kratochwill, T. R., & Gresham, F. M. (2007). Proactive, early screening to detect behaviorally at-risk students: Issues, approaches, emerging innovations, and professional practices. *Journal of School Psychology, 45*, 193-223.
- Simonsen, B., Myers, D., & Briere III, D. E. (2011). Comparing a behavioral check-in/check-out (CICO) intervention to standard practice in an urban middle school setting using an experimental group design. *Journal of Positive Behavior Interventions, 13*, 31-48.
- Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 24*, 323-355.
- Skiba, R. J. (2002). Special education and school discipline: A precarious balance. *Behavioral Disorders, 27*, 82-96.
- Skiba, R. J., Horner, R. H., Chung, C. G., Karega Rausch, M., May, S. L., & Tobin, T. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review, 40*(1), 85.
- Skiba, R. J. & Peterson, R. L. (2000). School discipline at a crossroads: From zero tolerance to early response. *Exceptional Children, 66*, 335-347.
- Skiba, R. J. & Peterson, R. L. (1999). The dark side of zero tolerance: Can punishment lead to safe schools? *Phi Delta Kappan, 80*, 372-382.
- Smith, B., & Fox, L. (2003). Systems of service delivery: A synthesis of evidence relevant to young children at risk of or who have challenging behavior. *Center for Evidence-based Practice: Young Children with Challenging Behavior*.

- Smith, S. C., Lewis, T. J., & Stormont, M. (2011). The effectiveness of two universal behavioral supports for children with externalizing behavior in Head Start classrooms. *Journal of Positive Behavior Interventions, 13*, 133-143. doi: 10.1177/1098300710379053
- Snell, M. E., Berlin, R. A., Voorhees, M. D., Stanton-Chapman, T. L., & Hadden, S. (2011). A survey of preschool staff concerning problem behavior and its prevention in Head Start classrooms. *Journal of Positive Behavior Interventions, 14*, 98-107.
- Sprague, J., Colvin, G., & Irvin, L. (1996). *The Oregon School Safety Survey*. Eugene: University of Oregon.
- Sprague, J., & Perkins, K. (2009). Direct and collateral effects of the First Step to Success program. *Journal of Positive Behavior Interventions, 11*, 208-221.
- Squires, J., Bricker, D. (2007). *An Activity-Based Approach to developing young children's social emotional competence*. Baltimore, Maryland: Paul H. Brookes Publishing.
- Steed, E. A., & Webb, M. L. (2012). The psychometric properties of the Preschool-Wide Evaluation Tool (PreSET). *Journal of Positive Behavior Interventions*.
- Stormont, M., Lewis, T. J., & Beckner, R. (2005). Positive behavior support systems: Applying key features in preschool settings. *Teaching exceptional children, 37*(6), 42-49.
- Stormont, M., Lewis, T. J., & Smith, S. C. (2005). Behavior support strategies in early childhood settings: Teachers' importance and feasibility ratings. *Journal of Positive Behavior Interventions, 7*, 131-139. doi:10.1177/10983007050070030201
- Stormont, M., Smith, S. C., & Lewis, T. L. (2007). Teacher implementation of precorrection and praise statements in Head Start classrooms as a component of a program-wide system of positive behavior support. *Journal of Behavioral Education, 16*, 280-290.

- Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M. et al. (2000). Applying positive behavior support and functional behavioral assessment in schools. *Journal of Positive Behavior Interventions*, 2, 131-143.
- Sugai, G., Horner, R. H., & Lewis-Palmer, T. (2002). *Universal team implementation checklists (2.2)*. Eugene: University of Oregon.
- Sugai, G. Lewis-Palmer, T., Todd, A., & Horner, R. H. (2001). *School-wide Evaluation Tool*. Eugene: University of Oregon.
- Sugai, G., Todd, A. W., & Horner, R. (2000). *Effective Behavior Support (EBS) Survey: Assessing and planning behavior supports in schools*. Eugene: University of Oregon.
- Tapp, J., Wehby, J., & Ellis, D. (1995). A multiple option observation system for experimental studies: MOOSES. *Behavioral Research Methods, Instruments, and Computers*, 27, 25-31.
- Walker, H. M., Horner, R. H., Sugai, G., Bullis, M., Sprague, J. R., Bricker, D., & Kaufman, M. J. (1996). Integrated approaches to preventing antisocial behavior patterns among school-age children and youth. *Journal of Emotional and Behavioral Disorders*, 4, 194-209.
- Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence: *Journal of Consulting and Clinical Psychology*, 61, 620.
- Webster-Stratton, C., & Hammond, M. (1997). Treating Children with Early-onset conduct problems: a comparison of child and parent training interventions. *Journal of consulting and clinical psychology*, 65(1), 93.

- Wenz-Gross, M. & Upshur, C. (2012). Implementing a primary prevention social skills intervention in urban preschools: Factors associated with quality and fidelity. *Early Education and Development, 23*, 427-450. doi: 10.1080/10409289.2011.589043
- Whitted, K. S. (2011). Understanding how social and emotional skill deficits contribute to school failure. *Preventing School Failure: Alternative Education for Children and Youth, 55*(1), 10-16.
- Wilder, D. A., Atwell, J., & Wine, B. (2006). The effects of varying levels of treatment integrity on child compliance during treatment with a three-step prompting procedure. *Journal of Applied Behavior Analysis, 39*, 369-373.
- Wittmer, D., Doll, B., & Strain, P. (1996). Social and emotional development in early childhood: The identification of competence and disabilities. *Journal of Early Intervention, 20*(4), 299-318.
- Zins, J. E., Bloodworth, M. R., Weissberg, R. P., Walberg, H. J. (2007). The scientific base linking social emotional learning to school success. *Journal of Educational and Psychological Consultation, 17*, 191-210. doi: 10.1080/10474410701413145.

Table 1

Program and Classroom Characteristics school year 2010-2011

Preschool (n=10)	Program type	# of classes (n=58)	# of children in program	Teacher to child ratio in program	% children in program with IFSPs	% minority	% male/female
Intervention 1	Private	14	150	1:8-10	10.82	3.1	55/45
Intervention 2	Private	7	100	1:8-10	8.08	16.0	46/54
Intervention 3	Head Start	4	72	2:15-20	26.47	85.0	65/35
Intervention 4	Head Start	2	108	2:15-20	4.62	80.0	54/46
Intervention 5	Private	4	119	1:8-10	20.00	14.28	58/42
Intervention 6	Private	3	133	1:8-10	21.42	39.39	60/40
Control 1	Head Start	11	324	2:18	15.74	82	48/52
Control 2	Head Start	6	Not available	2:18	Not available	Not available	Not available
Control 3	Head Start	6	142	2:18	7.5	66	59/41
Control 4	Head Start	1	36	2:18	19	36	50/50

Note. Individualized Family Support Plan (IFSP)

Table 2

Original versus transformed assumptions

Original Assumptions	Control TPF	Intervention TPF	Control TBC	Intervention TBC
Skewness	.559 Met	.494 Met	5.146 Violated	2.553 Violated
Kurtosis	.361 Met	.432 Met	46.628 Violated	16.027 Violated
Homogeneity of Variance	$p < .001$ Violated	$p < .001$ Violated	$p < .001$ Violated	$p < .001$ Violated
Transformed Assumptions	Control TPF	Intervention TPF	Control TBC	Intervention TBC
Skewness	-1.31 Met	-1.295 Met	-.229 Met	-.582 Met
Kurtosis	2.877 Met	2.931 Met	-.582 Met	.780 Met
Homogeneity of Variance	$p < .001$ Violated	$p < .001$ Violated	$p = .093$ Met	$p = .104$ Met

Table 3

Descriptive Statistics for Mixed ANOVA Total Protective Factors (TPF) & Total Behavior Concern (TBC)

Measure	Condition	N	Test Score Data	
			Mean	SD
Pre TPF	Control	394	69.93	14.39
	Intervention	309	74.03	10.64
Post TPF	Control	394	76.68	15.30
	Intervention	309	75.01	9.98
Total		703		
Measure	Condition	N	Mean	SD
Pre TBC	Control	394	10.30	9.27
	Intervention	309	7.44	4.39
Post TBC	Control	394	9.14	6.75
	Intervention	309	6.61	3.87
Total		703		

Table 4

Descriptive Statistics for HLM Levels 1 & 2

Levels	Variable Name	N	Test Score Data			
			<i>M</i>	<i>SD</i>	Min	Max
Level 1	Pre TPF	349	74.17	10.68	34	102
	Post TPF	349	75.73	10.32	42	102
	Pre TBC	349	6.78	3.88	0	18
	Post TBC	349	6.78	3.88	0	18
Level 2	Treatment Fidelity	21	67.29	12.21	35	82
	Years of Experience	21	11.48	7.28	1	24
	Education Level	21	.81	.40	0	1

Note. TPF=Total Protective Factor; TBC=Total Behavior Concerns

Table 5

Summary Statistics for the Final Multilevel Model (random intercept only) for TPF

Fixed Effect Parameter (with robust standard errors)	Coefficient	Standard error	t	df	p
For Intercept 1, β_0					
Intercept 2, γ_{00}	43.171	17.484	2.469	17	0.024
TF, γ_{01}	0.188	0.243	0.778	17	0.447
Years of Experience, γ_{02}	-1.121	0.376	-2.977	17	0.008
Education Level, γ_{03}	-3.558	6.356	-0.560	17	0.583
For Pre-TPF slope, β_1					
Intercept 2, γ_{10}	0.519	0.201	2.580	324	0.010
TF slope, γ_{11}	-0.004	0.003	-1.255	324	0.210
Years of Experience, γ_{12}	0.016	0.006	2.797	324	0.005
Education Level, γ_{13}	0.034	0.091	0.378	324	0.706
Random Effect	Standard Deviation	Variance	df	X^2	p
Intercept, u_0	4.775	22.807	17	131.509	<0.001
Level 1, r	7.43193	55.233			

Table 6

Summary Statistics for the Final Multilevel Model (random intercept only) for Total Behavior

Concerns

Fixed Effect Parameter (with robust standard errors)	Coefficient	Standard error	<i>t</i>	<i>d.f.</i>	<i>p</i>
For Intercept 1, β_0					
Intercept 2, γ_{00}	5.506	3.337	1.650	17	0.117
TF, γ_{01}	-0.007	0.050	-0.149	17	0.883
Years of Experience, γ_{02}	-0.186	0.064	-2.886	17	0.010
Education Level, γ_{03}	0.801	0.817	0.981	17	0.340
For Pre-TBC slope, β_1					
Intercept 2, γ_{10}	0.258	0.311	0.832	324	0.406
TF slope, γ_{11}	0.001	0.005	0.155	324	0.877
Years of Experience, γ_{12}	0.017	0.007	2.433	324	0.016
Education Level, γ_{13}	-0.084	0.147	-0.572	324	0.568
Random Effect	Standard Deviation	Variance	<i>d.f.</i>	χ^2	<i>p</i>
Intercept, u_0	1.508	2.267	17	91.828	<0.001
Level 1, r	2.852	8.136			

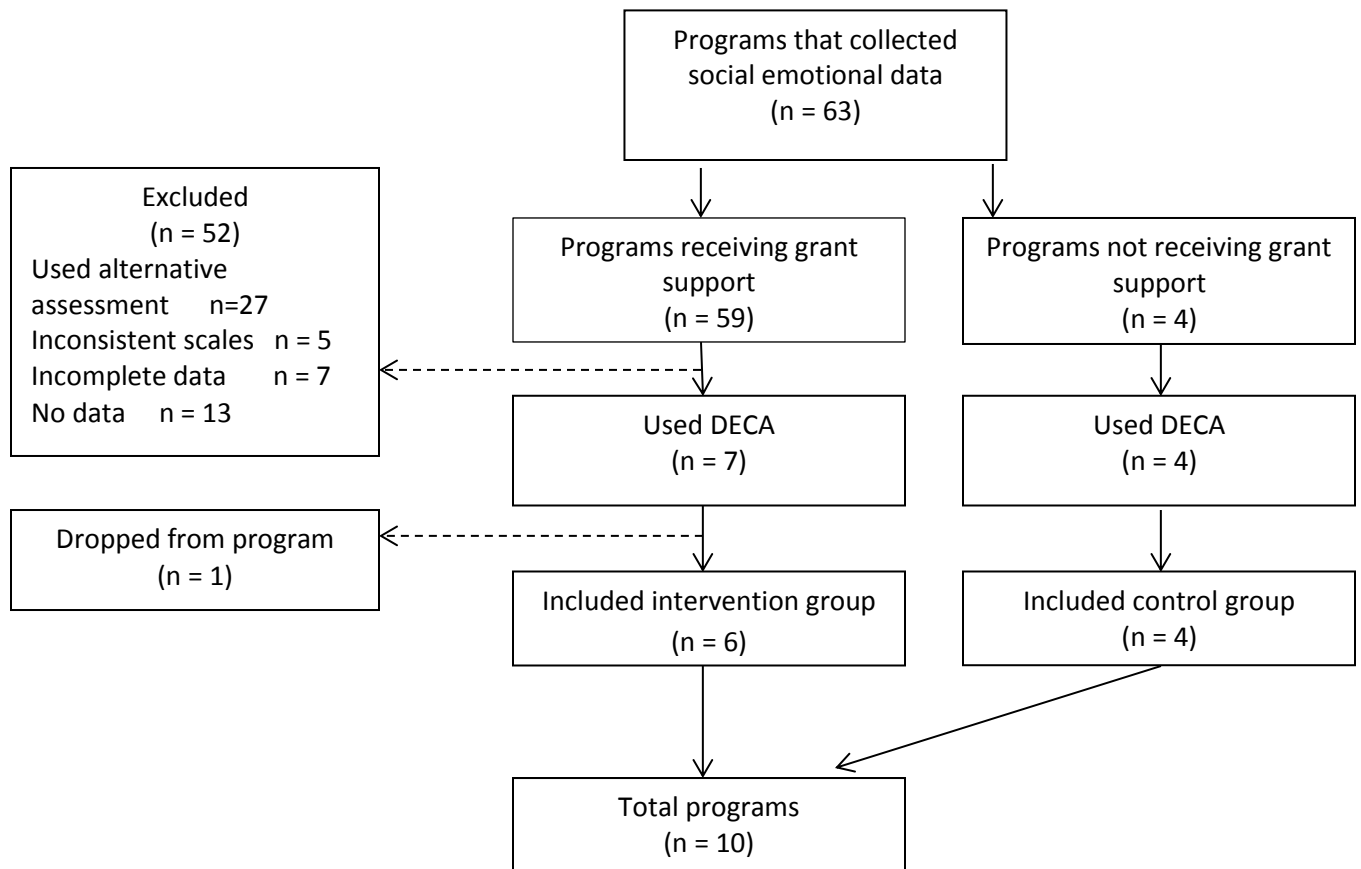


Figure 1. Flowchart of participating programs. Devereux Early Childhood Assessment (DECA; LeBuffe & Naglieri, 1999). Inconsistent scales indicates programs that used one of the rating scales at pre-test and a different rating scale at post-test. Incomplete data indicates when data from only one (pre or post) were reported.

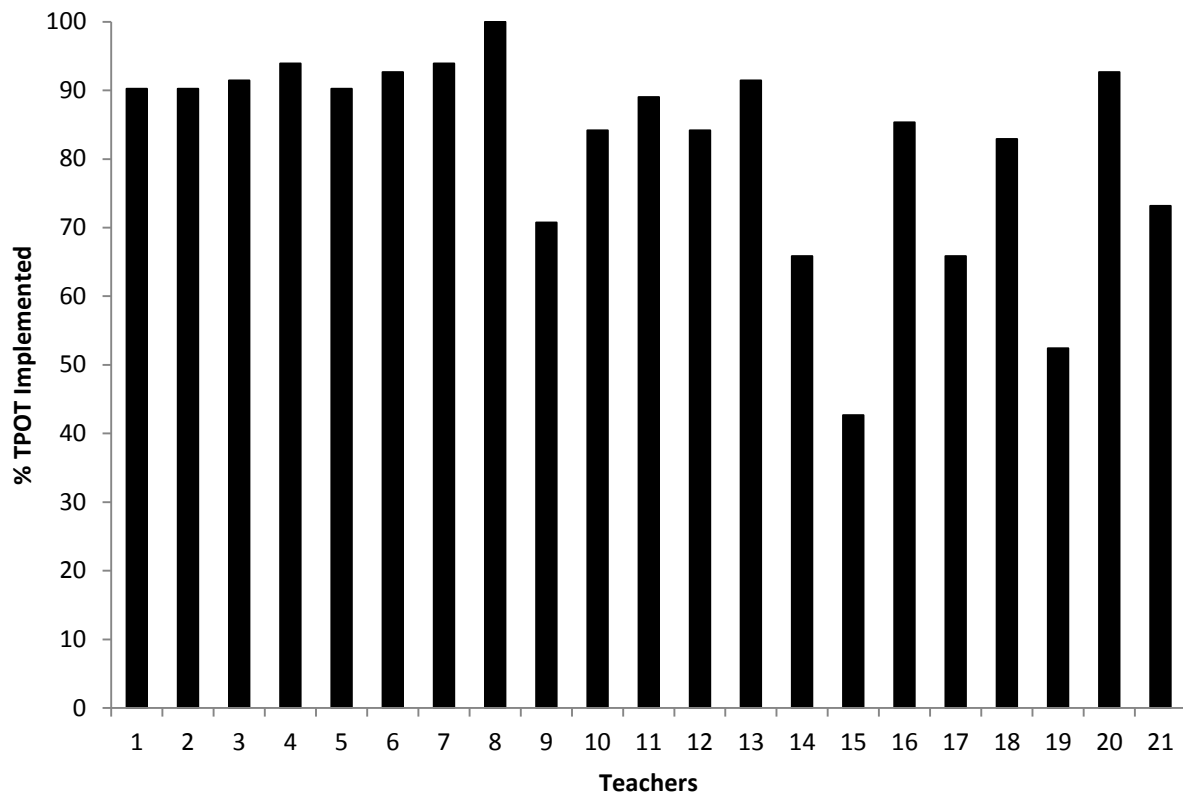


Figure 2. Percentage of items on the TPOT implemented with fidelity for 21 teachers across 6 programs implementing PWPBIS.

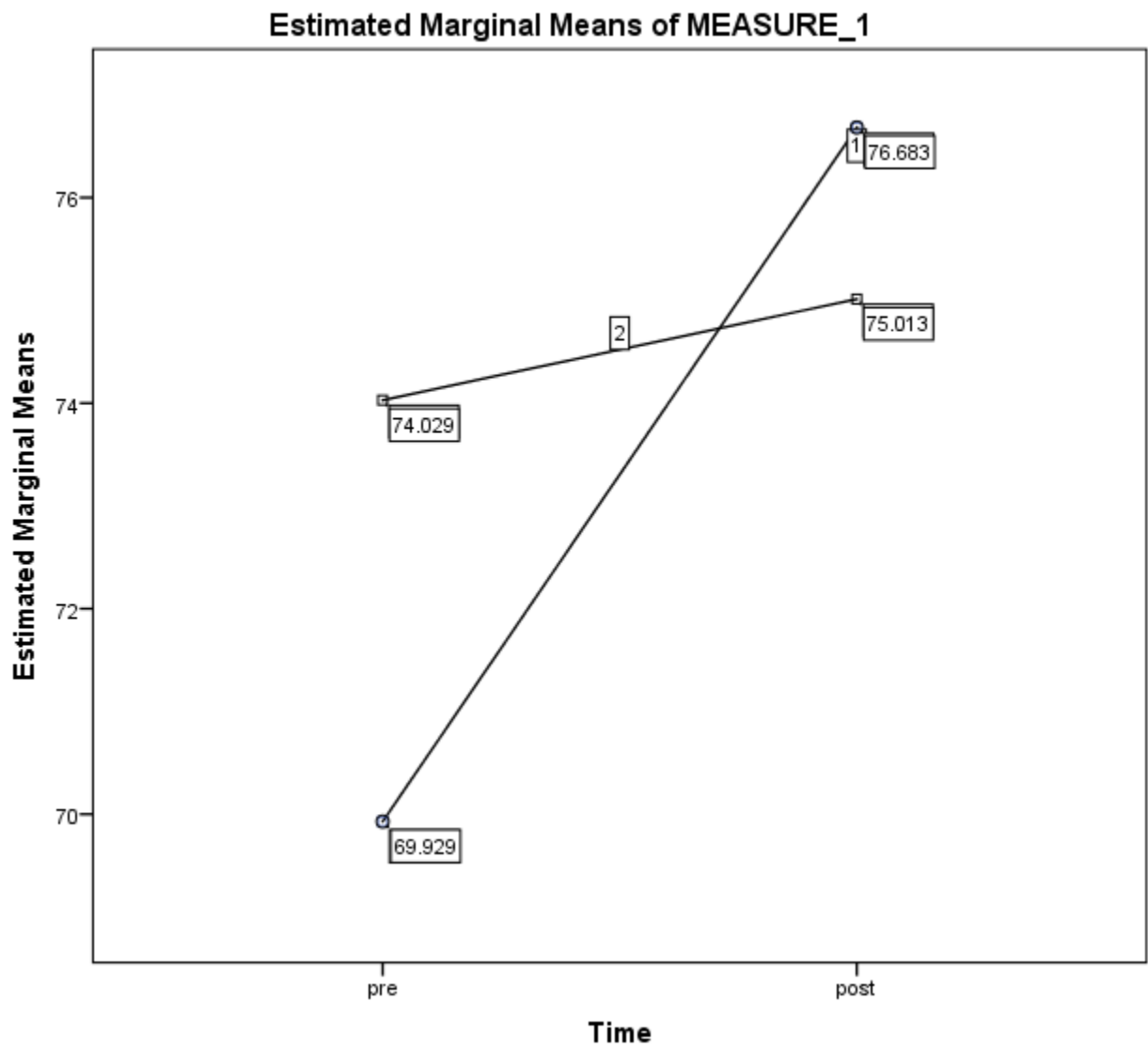


Figure 3. Profile Plot for Mixed ANOVA TPF; 1= control group; 2= intervention group

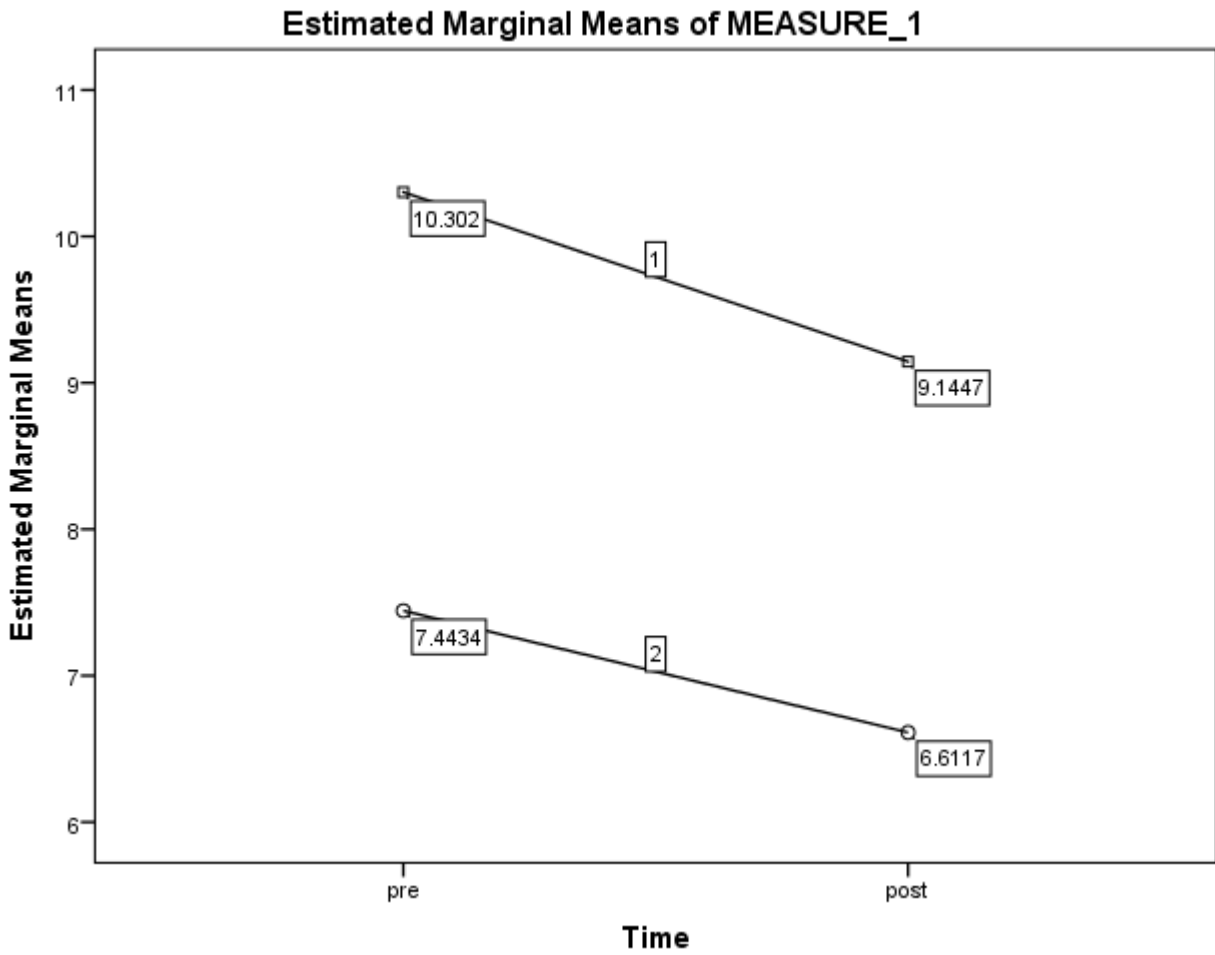


Figure 4. Profile Plot for Mixed ANOVA TBC; 1= control group; 2= intervention

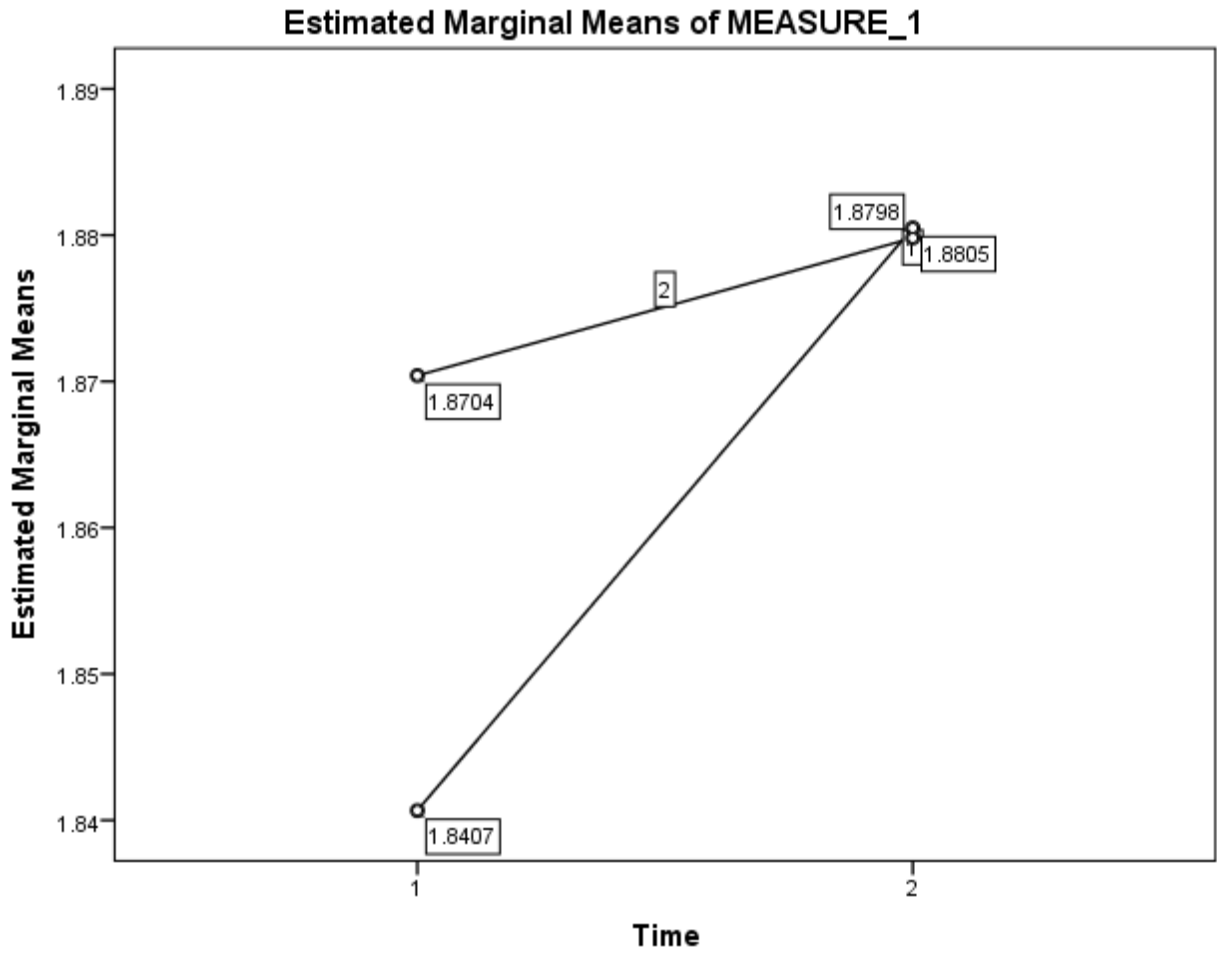


Figure 5. Profile Plot for transformed Mixed ANOVA TPF; 1= control group; 2= intervention group

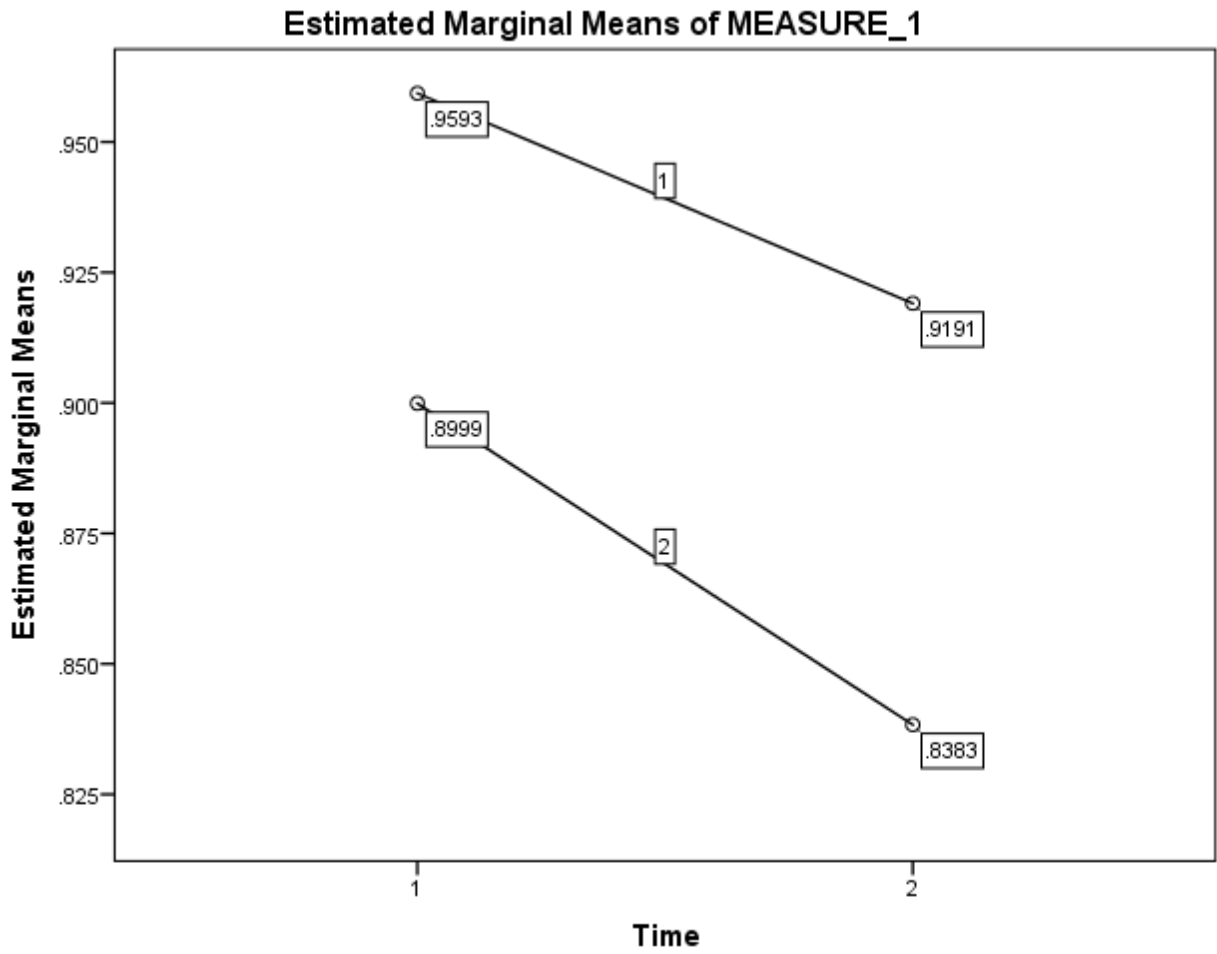


Figure 6. Profile Plot for transformed Mixed ANOVA TBC; 1= control group; 2= intervention

Curriculum Vita

August 2013

Robin Drogan

Education

2008-present	Lehigh University, Special Education
Doctoral Student	Bethlehem, Pennsylvania
	<i>Ph.D. anticipated September 2013</i>
2006-2008	Bloomsburg University, Special Education
Masters of Science	Bloomsburg, Pennsylvania
1988-1993	University of Maryland, Special Education
Bachelors of Science	College Park, Maryland

Certification

Pennsylvania Teacher Certification in Special Education: Level II

Professional Experiences

	Central Susquehanna Intermediate Unit, Montandon, Pennsylvania
2001-2008	Teacher of Special Education; Itinerant Autistic Support, Classroom Autistic Support, Classroom Multiple Disabilities Support
	Project All Children Together, A Division of Abilities Network, Towson, Maryland
2000-2001	Project ACT Teacher; Behavioral Support Teacher for Early Childhood Programs

Howard County Public School System, Ellicott City, Maryland

1993-2000

Teacher of Special Education;

Classroom Multiple Disabilities, Itinerant Work Study,

Co-Teacher grades K-3

Teaching Experience

Lehigh University

TLT/SPED 411 - Early Childhood Education,

Co-Instructor with Suzanne Mulhern, M.Ed.

SPCH/SPED 202 - Applied Behavior Analysis

Co-Instructor with Christine Cole, PhD

TLT/SPED 422 - Student Teaching Supervision

Co-Instructor with Lee Kern, PhD

SPED 428 – Positive Behavior Support

Co-Instructor with Lee Kern, PhD

SPED 332 - Education and Inclusion for Individuals with Special Needs,

Co-Instructor with Mary Beth Calhoon, PhD

SPED 452 - Assessment in Special Education,

Co-Instructor with Minyi Shih, PhD

SPED 465 - Advanced Methods of Inclusion

Co-Instructor with Natalie Sokol, PhD

SPED 330 – Special Topics in Special Education: Autism

Co-Developer and Co-Instructor with Brenna Wood, PhD

Research Experience

Fall 2012-present <i>Senior Facilitator</i>	Promoting Engagement with ADHD PreKindergarteners (<i>Project PEAK</i>), A study to further develop and refine family education sessions to increase engagement and implementation of evidence based practices by parents of young children at risk for ADHD. Investigators, Dr. George DuPaul (PI) and Dr. Lee Kern, Lehigh University.
Fall 2011-Spring 2012 <i>Research Assistant</i>	Autism Speaks Grant, Lehigh University A pilot study to investigate peer mediated intervention to support social conversation of high school students with autism. Investigators, Dr. Linda Bambara (PI) and Dr. Christine Cole (Co-PI), Lehigh University.
Fall 2009-Spring 2010 <i>Graduate Assistant</i>	A study implementing a tier 2 math intervention to 2 nd 2 nd grade students. Investigator, Dr. Minyi Shih. Collaboration in functional behavior assessment and function-based intervention in an early childhood setting with Dr. Brenna Wood.
Fall 2008 <i>Research Assistant</i>	Center for Adolescent Research in Schools (<i>CARS</i>), A study to evaluate intervention for adolescents with emotional and behavioral disorders. Principal investigator, Dr. Lee Kern, Lehigh University.

Fall 2007	Dr. Sheila Jones
Spring 2008	Dr. Sheila Jones and Dr. Darlene Perner
<i>Graduate Assistant</i>	Exceptionalities Department, Bloomsburg University

Awards and Recognition

2012 *Ted Carr Outstanding Poster Presentation Award*, Honorable Mention
Drogan, R. R., & Kern, L. Effects of a tier 2 intervention for young children
Presented as a poster at the International Conference on Positive Behavior
Support, Atlanta, GA.

Journal Articles

Drogan, R. R. & Kern L. (in press). Examination of the mechanisms underlying effectiveness of the Turtle Technique. *Topics in Early Childhood Special Education*.
Wood, B. K., Drogan, R. R. & Janney, D. (Accepted pending revisions). Early childhood teacher involvement in functional behavior assessment and function-based interventions: A literature review.

Professional Presentations

Wood, B. K., Drogan, R. R., & Janney, D. M. (2013, March). *Early childhood teacher involvement in functional behavioral assessment and function-based interventions*. Poster accepted for presentation at the 10th International Conference on Positive Behavior Support, San Diego, CA.
Goh, A., & Drogan R. (2012, December). *Function-based interventions for stereotypic and self-injurious behaviors in schools*. Paper presented at TASH Conference, Long Beach, CA.

Bambara, L. B., Cole, C., Kunsch, C., Drogan, R., & Tsai, S. (2012, November). *Teaching peers to support conversational skills of high school students with autism*. Paper presented at TASH Conference, Long Beach, CA.

Drogan, R., & Kern, L. (2012, October). *The Turtle Technique: The application of a tier 2 behavioral intervention with young children*. Paper presented at 36th Annual National Teacher Educators for Children with Behavior Disorders Convention on Severe Behavior Disorders of Children and Youth, Tempe, AZ.

Drogan, R., & Wood, B. K. (2012, October). *Early childhood teachers and young children at risk for emotional and behavioral disorders*. Paper presented at the 36th Annual National Teacher Educators for Children with Behavior Disorders Convention on Severe Behavior Disorders of Children and Youth, Tempe, AZ.

Wood, B. K., & Drogan, R. (2011, October). *Teacher Use of FBA and Function Based Interventions in Early Childhood Settings*. Paper presented at the 36th Annual National Teacher Educators for Children with Behavior Disorders Convention on Severe Behavior Disorders of Children and Youth, Tempe, AZ.

Goh, A., & Drogan, R. (2012, April). *Linking Intervention to Function of Problem Behavior*. Poster presented at the International Council for Exceptional Children Conference, Denver, CO.

Shih, M., & Drogan, R. (2012, April). *Efficacy of a Tier II Mathematics Intervention for Second Graders*. Paper presented at the International Council for Exceptional Children Conference, Denver, CO.

Drogan, R. R., & Kern, L. (2012, March). *Effects of a tier 2 intervention for young children*. Poster presented at the 9th International Conference on Positive Behavior Support, Atlanta, GA.

Goh, A., & Drogan, R. (2011, April). *The Application and Effectiveness of School-based individualized PBS with Problem Behaviors: 1997-2009*. Paper presented at the International Council for Exceptional Children Convention, Washington, DC.

Singley, D., & Drogan, R. (2011, April) *Improving Independent Living Skills with an Adult with Autism*. Poster presented at the International Council for Exceptional Children Conference, Washington, DC.

Knoster, T., Llewelly, G., & Drogan, R. (2008, March). *Team Functional Assessment Interview: The Initial Line of Inquiry*. Poster presented at the 5th Annual International Conference on Positive Behavior Support, Chicago, IL.

Service

Manuscript Reviews

2012	Topics of Early Childhood Special Education, Co-Reviewer with Brenna Wood, PhD
2012	Journal of Behavioral Education, Guest Reviewer
2012	Journal of Behavioral Education, Co-Reviewer with Lee Kern, PhD

Professional

2012-present	Association for Positive Behavior Support Student Network Leadership Committee Dissemination Committee
--------------	--

Community

2010-2011	Behavior Consultant, SPARK Early Childhood, Bethlehem Area School District, Bethlehem,
-----------	---

Pennsylvania

Professional Membership

Council for Exceptional Children (CEC)

Association for Positive Behavior Support (APBS)