UNIVERSITY OF CALIFORNIA
RIVERSIDE

# The Use of Performance Feedback to Improve Treatment Integrity and Student Behavior With a Class-Wide Behavior Intervention

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Education

by

Diana Ginns Socie

June 2016

Dissertation Committee:
Dr. Austin Johnson, Chairperson
Dr. Cathleen Geraghty
Dr. Marsha Ing

The Dissertation of Diana Ginns Socie is approved:

_____

_____

_____
Committee Chairperson

University of California, Riverside

ABSTRACT OF THE DISSERTATION

The Use of Performance Feedback to Improve Treatment Integrity and Student Behavior With a Class-Wide Behavior Intervention

by

Diana Ginns Socie

Doctor of Philosophy, Graduate Program in Education
University of California, Riverside, June 2016
Dr. Austin Johnson, Chairperson

Students with emotional disturbance (ED) who receive special education services in a public school setting fare significantly worse than other students; they are more likely to drop out of school and experience difficulties later in life than students who receive special education services under any other category. Research examining traditional attempts to provide staff with professional development to improve their ability to work successfully with these students has shown that these efforts typically produce unsuccessful outcomes. As a result, this study includes the development of a class-wide behavior intervention based on principles of Applied Behavior Analysis (ABA) and the use of performance feedback to support the implementation of the intervention. Using a multiple baseline across classrooms design, the effects of performance feedback on the treatment integrity of the system, along with subsequent changes in student engagement and student disruptive behavior, were measured. Results indicated that the implementation of performance feedback was associated with increases in staff treatment integrity. Additionally, improvements in student academic engagement and reductions in disruptive behavior

were observed. Limitations of the study, implications for practice, and directions for future research are discussed.

Table of Contents

List of Tables

List of Figures

List of Appendices

# Chapter 1: Introduction

The number of students in grades kindergarten through 12th grade eligible for special education with emotional disturbance (ED) has been steadily increasing since Congress enacted the Education for All Handicapped Children Act in 1975 (PL 94-142; now known as the Individuals with Disabilities Education Act [IDEA, 2004]); the number of students identified with emotional disturbance rose from 283,000 in 1976 to 407,000 in 2009 nationwide, a nearly 44% increase (U.S. Department of Education [USDOE], 2011). This increase in the number of students identified for special education is unique to this eligibility category; in comparison, during this same time period, the number of students eligible for special education with an intellectual disability has decreased by 53% (USDOE, 2011). This situation is complicated by the fact that compared to other students who receive special education services, students with ED demonstrate significantly more social and/or behavioral problems. These social and behavioral problems result in deleterious outcomes for these students, rendering them more likely to drop out of school, become incarcerated at an early age, and experience difficulty becoming gainfully employed after high school (Bradley, Doolittle, & Bartolotta, 2008). Unfortunately, this situation is further complicated by the fact that teachers for students with ED tend to be in short supply and unprepared to take on the challenge of working with these students (Henderson, Klein, Gonzalez, & Bradley, 2005).

## ED Student Characteristics

Although there is much debate about the current eligibility criteria for emotional disturbance (ED) under the Individuals with Disabilities Education Act (IDEA, 2004), the federal definition has remained essentially unchanged over the past 40 years (Merrell & Walker, 2004). Due to the broad nature of the ED eligibility criteria, students with a range of disorders, including mood disorders, anxiety, conduct disorder, oppositional defiant disorder, personality disorders, and/or attention deficit hyperactivity disorder who manifest concurrent academic deficits may qualify for special education services with ED (Cullinan, 2004). Researchers hypothesize that students with these types of mental health diagnoses exist along a continuum, where students at the low end experience little to no difficulty, and students on the high end of the continuum exhibit extreme behaviors under normal circumstances (Cullinan, 2004). In this vein, students at the high end of the continuum who demonstrate significant behavioral difficulties are less likely than their peers to achieve grade level standards (Sutherland, Lewis-Palmer, Stichter & Morgan, 2008).

To qualify for special education in the United States with an emotional disturbance, a student must demonstrate deficits in three key areas: (a) behavior, with excesses in aggression and disruptive classroom behavior and/or deficits, manifesting in social withdrawal and noncompliance; (b) academic learning problems resulting in lower achievement, and behavioral difficulties in maintaining attention; and (c) interpersonal relationships, which affect their social and language skill development (Landrum, Tankersley, & Kauffman, 2003). In general, students eligible with ED typically

demonstrate academic performance one grade-level or more behind their peers (Kauffman, 2001). Researchers have questioned the direction of this relationship between behavior and academics, hypothesizing that it is just as likely that academic underperformance results from behavioral problems as it is that academic difficulties cause behavioral disorders (Sutherland et al., 2008). At this time, it is safe to say the extent to which students' academic difficulties result in behavioral challenges is unclear (Sutherland et al., 2008).

Nonetheless, students with ED have often been categorized dichotomously as demonstrating either externalizing or internalizing behaviors (Furlong, Morrison, & Jimerson, 2004; Gresham & Kern, 2004). Students with externalizing behaviors display aggressive, argumentative, impulsive, and noncompliant behavior characterized by an "undercontrolled, acting-out style of responding" (Gresham, Lane, MacMillan, & Bocian, 1999, p. 231). Due to the nature of these externalizing problems, teachers are much more likely to report strong negative reactions to these students (Gresham et al., 1999; Sutherland & Oswald, 2005). Eventually, these types of negative interaction patterns can lead to transactional processes, where the students' noncompliant or aggressive behaviors render the teacher less likely to interact with him or her in a supportive fashion or avoid the student altogether (Sutherland & Oswald, 2005). Further, teachers are also more likely to refer students with externalizing behaviors to school administrators for disciplinary action and suspension (Gresham & Kern, 2004). Conversely, students with internalizing behaviors may also qualify for special education services under ED criteria, and are equally at risk for issues with long term psychological maladjustment resulting

from extreme withdrawal, depression, and/or suicidal ideation (Gresham et al., 1998). In addition to students with externalizing and internalizing disorders, it is important to note that there is evidence to support a third category of students who demonstrate comorbid externalizing and internalizing behavior problems (Gresham et al., 1998). Regardless, students in all categories are more likely to demonstrate poor social skills, be less accepted by their peers, and report being lonelier than other students throughout their school career (Gresham et al., 1998).

**Student Outcomes**

Using data from two longitudinal studies, the Special Education Elementary Longitudinal Study (SEELS) and the National Longitudinal Transition Study-2 (NLTS2), Bradley and colleagues (2008) found that less than half of students in ED programs received behavioral intervention services at any point in their educational career. Coupled with the fact that these students often lack the behavioral and social skills necessary to experience academic and vocational success, it is unsurprising that these students often face dismal outcomes; students with ED tend to be (a) involved in fights over two times as often as students with other disabilities, (b) expelled more than three times as often as students with other disabilities, (c) more likely to engage in substance abuse, (d) more than two times likely to drop out of school than general education students, and (e) significantly less likely than students with other disabilities to obtain employment after graduating high school (Bulis & Cheney, 1999; Merrell & Walker, 2004; Bradley et al., 2008). Finally, although recent studies have produced mixed results with respect to youth incarceration rates for adolescents with ED, a conservative estimate for the number of

adolescents in juvenile detention centers with ED is 32% (Quinn et al., 2001), while other estimates are as high as 47.7% (Quinn et al., 2005).

Without effective intervention to remediate these deficits, the behavior and academic deficiencies students with ED manifest as early as kindergarten often continue throughout their lives (Quinn & Poirier, 2004). Additionally, without intervention, many of these students develop maladaptive behavior patterns resulting in delinquency, criminal activity, and poor health outcomes later in life (Cohen, Piquero, & Jennings, 2010). As a result, the costs of supporting a student with ED can be staggering; analyses of long-term outcomes have revealed there are myriad financial repercussions to taxpayers and the public, along with various social costs that reduce the aggregate well-being of society (Quinn & Poirier, 2004), including dependence on the welfare system, motor vehicle accidents, lack of productivity, strain on the foster care system, and child abuse or neglect (Cohen et al., 2010). Additionally, with respect to personal outcomes, students with ED face numerous opportunity costs and private costs, such as long periods of unemployment, medical costs, and teen pregnancy, that serve to diminish their overall emotional well-being and contribution to society throughout their lives (Quinn & Poirier, 2004; Cohen, et al. , 2010).

Although not all students with ED experience these kinds of outcomes, research suggests that a substantial number do. Results from the National Longitudinal Transition Study-2 (NLTS-2; Wagner, Kutash, Duchnowski, Epstein, & Sumi, 2005) indicate that 44% of students with ED drop out of high school, 11% of males with ED ages 15-19 have fathered a child, and, of those students surveyed, 43% have been on probation or parole.

Further, a recent analysis from the U.S. Department of Education indicates that the number of students with ED who have dropped out of high school is significantly greater than students from any other disability category (USDOE, 2014).

In the most comprehensive survey of incarcerated youth with disabilities, Quinn and colleagues (2005) determined that on average, 33% of juveniles in detention centers are or have been recipients of special education services. Of those, the highest percentage (47.7%) were students identified with ED (Quinn, Rutherford, Leone, Osher, & Poirer, 2005). With respect to the costs associated with these students engaging in criminal behavior, depending on when a particular at-risk youth begins to interface with the justice system (e.g., jail, police, court system), Cohen and Piquero (2009) estimate the cost to society for juvenile delinquency is between $2.6 and $5.3 million. The cost estimates vary depending on which type of activities the youth engages in, and at what age. For example, depending on when they drop out of school, at-risk youth may cost the system between $300,000 and $450,000 (Cohen & Piquero, 2009). Additionally, youth involvement in drugs and alcohol may run as high as $740,000 for a heavy drug user (Cohen & Piquero, 2009). Further, teen pregnancy, from medical costs for pregnant teens, to the cost of raising the children of teen mothers costs taxpayers an aggregate of $9.1 billion. Lamentably, children of teenage mothers also face their own set risks, which continue to impact the overall well-being of society at large. For example, children of teen mothers are 10.6% more likely to become incarcerated at some point in their lives (Cohen et al., 2010).

Given the social and financial costs associated with juvenile justice in general and ED in particular, there appear to be various financial repercussions for not intervening early on to prevent or remediate children's behavioral and mental health problems. Researchers have continued to echo this sentiment, indicating that schools that implement evidence-based programs earlier in students' school careers have the potential to produce significant change in student trajectories (Cohen, Piquero, & Jennings, 2010, p. 424). Further, Cohen and colleagues (2010) relate that early interventions are a significantly effective, cost-conscious alternative to the primarily reactive programs (e.g., prison) currently used to curtail behavior problems when these at-risk children become adults. In a recent meta-analysis, Cohen and colleagues (2010) determined "early childhood intervention [evidence-based interventions implemented before 5th grade] is successful in preventing/reducing behavior problems, including crime and delinquency" (p. 398). However, despite the financial, educational, and health related advantages to providing early intervention, only approximately 3-9% of monies spent on dealing with these problems are funneled towards preventative efforts (Quinn & Poirier, 2004).

**Staff Training for ED Classrooms**

In many districts, due to the nature and severity of behavior problems of students, students identified with ED spend the majority of their school day outside of the general education setting, oftentimes in self-contained classrooms (USDOE, 2013). Based on data from the 2008-2009 school year, the most recent year for which this information is available, 40.6% of students with ED spent 80% or more of the school day outside of the general education setting (USDOE, 2013). This percentage is not including the 13.2% of

students with ED who were placed in separate schools for students with disabilities (USDOE, 2013). Moreover, the USDOE reports 18.8% of students with ED spend 40-79% of their day outside of the general education setting, and only interface with their peers in general education during physical education class or lunchtime (USDOE, 2013; Lane, Wehby, Little, Cooley, 2005).

Overall, what is clear from these data is that the majority of students with ED receive most of their educational programming in self-contained environments, in a setting away from general education students (USDOE, 2013). Although these self-contained classrooms have the potential to be advantageous to students with ED, teachers in these classrooms often do not use specific behavioral and instructional strategies to increase student learning opportunities (Maggin, Wehby, Moore Partin, Robertson, & Oliver, 2011). Additionally, the majority of teachers in ED classrooms lack the specific training necessary to intervene and provide appropriate behavior management for their students (Oliver & Reschly, 2010).

**Teacher Training**. Researchers have emphasized that one of the most salient barriers to educating children with ED is lack of teacher training (Brunsting, Sreckovic, & Lane, 2014; Cancio & Johnson, 2007). By law, teachers who work with students with ED in self-contained classrooms are required to hold a special education teaching credential (Oliver & Reschly, 2010). Unfortunately, due to teacher shortages, teachers of ED classrooms are less likely to have master's degrees and more likely to be working with an emergency or temporary credential than other special education teachers (Henderson et al., 2005; Bradley et al., 2008). In fact, a recent survey of 859 teachers of

students with ED revealed that when compared to other special education teachers, teachers in ED classrooms evidenced fewer years of classroom experience (mean of 12.0 compared to 14.7 years for non-ED teachers) and lacked credentials (e.g., they were less likely to hold a master's degree, more likely to be on an emergency credential only, and less likely to be fully certified for their assignment; Henderson et al., 2005). Further, special education teachers who work in ED classrooms rarely receive behavioral training specific to the needs of the students they work with (Brunsting et al., 2014; Begeny & Martens, 2006), and in one survey, only approximately a quarter of teachers had received at least an eight hour district-level training on working with students with disabilities (Bradley et al., 2008).

In addition, it appears that even teachers who attend credentialing programs are significantly underprepared to teach students with ED. In a recent study, Oliver and Reschly (2010) surveyed university teacher credentialing programs, analyzing course content for amount and quality of behavior management courses offered. They found that most universities provided some form of instruction around behavior management, but that most of this content was limited in scope and, when available, only embedded in other classes about teaching. As a result, special educators-in-training lacked exposure to sufficient learning opportunities to understand and intervene in students' problem behavior.

To help assess the extent to which teachers in training programs receive the necessary prerequisite training to work with students with behavioral difficulties, Begeny and Martens (2006) sampled 110 graduate students in master's programs for elementary,

secondary, or special education. Results of their study indicate that these graduate students received very little training in behavioral assessment, instruction, and intervention. Additionally, they found that special education teachers-in-training tended to lack sufficient training in working with children with behavior difficulties and applying behavioral strategies in the classroom (Begeny & Martens, 2006).

This issue is further complicated by research suggesting that on the whole, teachers who work with students with ED face more stressful situations on a daily basis than other special education teachers (Baker, 2005; Cancio & Johnson, 2007). This has led many researchers to hypothesize that the types of behavior problems typically found in ED classrooms (e.g., aggression towards staff, aggression towards other students, noncompliance, property destruction) can become aversive for teachers, which may subsequently cause teachers to interact less positively with students (Wehby, Symons, & Canale, 1998). As previously discussed, this pattern of negative interaction between teachers and students may create an environment in which students react to the negativity from the teacher and become more disruptive, rendering it more difficult for these students to gain the necessary social skills required to return to general education (Brunsting et al., 2014). As a result of these stressful situations and the cycle of negative interactions with students, ED teachers may experience more burnout and stress than other special education teachers (Brunsting et al., 2014). This phenomenon is associated with higher attrition rates and/or deterioration of staff members' physical and emotional health. Indeed, staff at the highest level of burnout can begin to struggle with depression,

demonstrate susceptibility to viruses, musculoskeletal pain, and chronic fatigue (Brunsting et al., 2014; Bradley et al., 2008).

Additionally, there are other negative outcomes associated with students' behavioral problems in ED classrooms. When teachers who lack sufficient training to work with students with ED are continuously exposed to students' engagement in aggressive and defiant behavior, one unintended consequence is that teachers may be less likely to implement behavioral interventions (Landrum et al., 2003; Oliver & Reschly, 2010). Moreover, teachers in these classrooms may be reluctant to implement empirically-based behavioral interventions, having formed opinions regarding the lack of efficacy of these interventions based on their experiences with incorrect or insufficient implementation (Cook, Landrum, Tankersley, & Kauffman, 2003). Regrettably, when teachers are misinformed as to the efficacy of behavioral interventions, it may lead teachers to question the value of interventions based on research in applied behavior analysis (Landrum et al., 2003). This problem appears to be pervasive, as many teachers of students with ED report low ability and willingness to use varied reinforcement schedules, continue to document students' response to interventions, and implement behavior intervention plans (Baker, 2005).

Each of these issues may contribute to teacher attrition, which presents its own set of difficulties for schools, children, and society. For example, when teachers leave a school, there are quantitative costs (e.g., costs incurred during recruitment and training of new teachers) and qualitative costs (e.g., decline of morale, organizational stability, negative impact on educational productivity; McLeskey & Billingsley, 2008). Currently,

there is no cost analysis available for the financial strain associated with teacher attrition specifically for students with ED; however, on average, approximately 6% of special education teachers leave the profession each year, and 22.75% of special education teachers do not continue in the same position from one year to the next (McLeskey & Billingsley, 2008). When considered together, these data suggest the cost of special education teacher attrition nationwide is in the billions of dollars (Darling-Hammond & Sykes, 2003).

Fortunately, it appears that when districts offer support for teachers, teacher attrition may be reduced (Brunsting et al., 2014). Although some of the support is related to financial incentives (e.g., salary), and therefore may not be within the purview of site administrators, some of the recommendations are feasible to implement at the school level (McLeskey & Billingsley; Lane, Jolivette, Conroy, Nelson & Benner, 2011). Some of these recommendations include providing more training to teachers to help improve their understanding of the disabilities of the students in their classrooms with emphasis on empirically-validated interventions for remediating problem behavior (Brunsting et al., 2014). Additionally, districts can provide training to school-site administrators to help teachers better manage some of the stress they face in relation to the student behaviors in their classrooms (Brunsting et al., 2014). Further, since special education teachers are involved in supervision and training of paraeducators in their classrooms, specific support in this area from site administrators can help to alleviate some of the burden that teachers face in this area (Brunsting et al., 2014).

With respect to teachers of ED students, in order to affect meaningful change, Lane and colleagues (2011) specifically recommend that school districts (a) ensure high fidelity of implementation of intervention practices, and (b) prepare teachers for the challenge of working with students with ED with improving the quality of professional development at school sites. Similarly, other researchers have proposed that improvement in the quality and process of professional development at the school level can impact teachers' sense of competency, their ability to delivery intervention with fidelity, their job satisfaction, their commitment to the profession, and their intent to stay in their current position (Billingsley, 2004; Gersten, Keating, Yovanoff, & Harniss, 2001).

**Paraprofessional Training**. In addition to teachers, many classrooms with ED students are also staffed with paraprofessionals. In general, the extent to which the support of paraprofessionals in special education settings helps to improve student outcomes is widely debated (Stockall, 2014). It appears that student outcomes are improved when paraprofessionals are well-trained and prepared for their role of supporting students and teachers (Hall, Grundon, Pope, & Romero, 2010); however, various issues influence the effectiveness of paraprofessional support, including the type of training received both outside of the district of hire (e.g., outside courses taken) and in the district (e.g., pre-service training and professional development), the relationship between the paraprofessional and the teacher, the ratio of paraprofessionals to students, and the type of opportunities in the district or school for ongoing support and professional development (Stockall, 2014; Giangreco, Suter, & Hurley, 2014). Unfortunately, due to lack of training, across the country, the majority of paraprofessionals working in special

education programs are underprepared to effectively support students, both behaviorally and academically (Giangreco, Edelman, Broer, & Doyle, 2001); although some states have developed training curricula for districts to follow when hiring paraprofessionals, most have not (Breton, 2010).

In general, paraprofessionals come to the classroom untrained or undertrained (Breton, 2010). Additionally, it appears that the least qualified paraprofessionals are often placed with the most challenging students, including students in ED classrooms (Breton, 2010). However, this lack of training does not appear to be due to lack of willingness on the part of these educators. In fact, when surveyed, the vast majority of paraprofessionals who lack training have expressed a continuing desire for more training and professional development (Breton, 2010). Moreover, the number one area in which the overwhelming majority of paraprofessionals feel they need more training is in "dealing with student behavior, emotional and social challenges" (Brenton, 2010, p. 41).

**Empirically-based Classroom Management Practices**

Evidence suggests that more structured classrooms are associated with more academically engaged students, and that students in more structured classrooms engaged in more socially appropriate behaviors (Simonsen et al., 2008). Key features in creating a structured classroom include setting up the physical arrangement of the room to minimize distractions (e.g., antecedent interventions), as well as posting, teaching, and reviewing expectations (Simonsen et al., 2008). Moreover, teacher delivery of contingent, specific praise, both within and independent of a group contingencies for classroom management (e.g., level systems, token economies, the good behavior game) are associated with

increases in student on-task and prosocial behavior (Simonsen et al., 2008; Cook et al., 2003).

Applying findings from research on classroom management to ED classrooms, Tankersley and colleagues (2004) suggest there are a variety of empirically supported behavioral interventions which may increase on-task, prosocial, and compliant behaviors. Since students with ED demonstrate both excesses in aggression and disruptive behavior with simultaneous deficits in social skills and compliance, researchers have hypothesized that class-wide interventions for these students should target both areas (Landrum, et al., 2003). Fortunately, many practices based on principles of Applied Behavior Analysis (ABA) have been shown to remediate the behavior problems of students with ED, targeting both behavioral excesses and deficits (Landurm et al., 2003). These strategies include reinforcement (positive, differential, negative), precision requests, behavioral momentum, time-out, response cost, group-oriented contingencies, and continuous monitoring of student performance (Landrum et al., 2003).

One research-based system used in ED classrooms that incorporates many of the ABA strategies previously discussed is a level system (Cancio & Johnson, 2007; Jones, Dohrn, & Dunn, 2004; Musser, Bray, Kehle, & Jenson, 2001; Mastropieri, Jenne, & Scruggs, 1998; Filcheck, McNeil, Greco, & Bernard, 2004). A level system includes the use of a level hierarchy in which students move up or down through the levels depending on the extent to which their behavior meets specific criteria (Cooper et al., 2007). Within a level system, appropriate behavior is shaped through the use of differential reinforcement of appropriate behaviors through praise and subsequent access to

performance-based level rewards. Additionally, the use of a level system requires "(a) a specific listing of expectations, requirements, and privileges associated with membership in each level, and (b) specific standards for either upward or downward mobility within the level system" (Mastropieri et al., 1998, p. 202).

Although there is research in the application of individual components of a level system (e.g., behavior specific praise) with students with ED, there is little research examining the overall impact of this type of system on student outcomes in educational settings (Mastropieri et al., 1998). Nonetheless, this type of system has the potential to be effective with ED students as it draws upon various principles from ABA, including antecedent manipulations of the environment, shaping, differential reinforcement, and sometimes a response cost element (Cancio & Johnson, 2007). Moreover, there are various advantages of using this type of system with students in ED classrooms, including providing a rigid structure with clearly delineated expectations (Jones et al., 2004; Musser et al., 2001).

In one of the only studies in which a level system was implemented with ED students, Mastropieri and colleagues (1998) were able to effectively decrease students' disruptive and off-task behaviors and increase students' task completion in a high school resource program. In this study, researchers worked with one special education teacher in a classroom with 15 high school students with behavioral disorders to implement a four-level color-coded intervention. Students earned levels based on their performance related to following the rules, engaging in work, and remaining in their seat during work time. One set of rules was common to all levels and then each level was associated with

different types of privileges. Employing a reversal design Mastropieri and colleagues (1998) concluded that the implementation of the intervention increased the number of assignments completed by students, the accuracy of completed work, and increased the amount of time students spent in their seats during the class period.

Despite the advantages of using this type of system, there are several barriers to its successful implementation (Cancio & Johnson, 2007). Many of these barriers are related to lack of support in implementation of the system, teachers' lack of training, and teachers' perceptions and attitudes regarding behavior change (Cancio & Johnson, 2007; Grieve, 2009).

**Barriers to Implementation**. Over the past 40 years, researchers have documented the effectiveness of class-wide interventions based on principles of applied behavior analysis. Nonetheless, for reasons discussed below, few teachers have successfully been able to adopt these principles into practice (Grieve, 2009). Some researchers explain this lack of implementation as a manifestation of teacher resistance, citing issues involving teacher beliefs (Piersel & Gutkin, 1983), level of intervention difficulty (Witt, 1986), teacher satisfaction with previous interventions (Witt, 1986; Grieve, 2009), perceived intervention effectiveness (Witt, 1986), teacher training and mechanism of teacher training and subsequent consultation (Piersel & Gutkin, 1983), and aversive teacher-student interactions that serve as punishment to teachers when engaging in positive teaching practices (Wheby, Simmons, Canale, & Go, 1998; Carr, Taylor & Robinson, 1991). However, the exact reason or reasons why this phenomenon continues

is unknown (Witt, 1986). Regardless, the reason appears to be multifaceted and complicated by the numerous factors discussed above (Grieve, 2009).

In an attempt to address these issues, Witt (1986) asked researchers to consider consumer (e.g., teacher, parent, student) satisfaction of different types of behavioral interventions. Witt argued that interventions rated as highly acceptable by teachers would be more likely to be continued after the consultant/researcher removed support. His most salient suggestion with respect to this issue was that interventions were products that need to be marketed (e.g., delivered) to consumers in a fashion similar to the business world to effectively produce adult and child behavior change. He emphasized that the discussion of effectiveness on the part of consultants is only one part of marketing, and he challenged researchers to discover other ways to effectively disseminate interventions.

One type of intervention for improving treatment integrity that may address some of the issues raised by Witt (1986) is performance feedback (PF). In the literature, PF has been described as a mechanism to insight adult behavior change whereby observers give systematic feedback to participants regarding their performance with respect to a specific criteria (Alvero, Bucklin, & Austin, 2001; Noell et al., 2005). Performance feedback will be discussed in more detail later, but it is important to note that the recipients of performance feedback consistently rated this type of intervention as highly acceptable (Alvero et al., 2001). Additionally, it also appears that because the mechanism by which this intervention functions varies (e.g., it may operate as an antecedent, a reinforcer, an establishing operation, or may serve multiple functions), PF has the advantage of affecting adult behavior through the application and management of various types of

behavioral contingencies (Alvero, Bucklin, & Austin, 2001). In fact, in a meta-analysis, Alvero and colleagues found seven different applications of PF, including feedback alone; feedback and antecedents; feedback and behavioral consequences; feedback and goal setting, feedback, antecedents, and behavioral consequences; feedback, antecedents goal setting, and behavioral consequences; and feedback, goal setting, and behavioral consequences.

**Class-wide interventions for ED in Secondary Settings.** As previously discussed, there are few studies evaluating the impact of research-based, class-wide interventions on student behavior in ED classrooms, and the majority of studies that have been published have evaluated outcomes in elementary classrooms. Nonetheless, similar to the recommendations of Simonsen and colleagues (2008), Cheney, Cumming, and Slemrod (2013) recommended that class-wide interventions at the secondary level incorporate the following components: physical organization of the classroom, teaching clear expectations and routines, the utilization of a structured system of reinforcement, and frequent praise provided to students. Similarly, Witt, VanDerHeyden, and Gilbertson (2004) provided recommendations based on ABA strategies for working with ED students derived from a comprehensive review of the literature for students at any level. These included teaching and reinforcing expected behaviors and consistent and accurate teacher responses to inappropriate student behavior.

**The Effect of Teacher Training on Intervention Implementation.** As previously discussed, on the whole, there appears to be a significant shortage of teachers who have the training and/or possess the knowledge and skills to adopt empirically-

based, class-wide behavioral interventions in the classroom (Tankersley et al., 2004). For example, with respect to the provision of behavior-specific praise, an effective and minimally-intrusive practice identified in the literature related to improvements in student outcomes in ED classrooms, most staff in ED classrooms use this strategy at an alarmingly low rate, and many times its rates are minimal in ED classrooms (Tankersley et al., 2004; Sutherland & Wehby, 2001; Sutherland, Wehby, & Copeland, 2000). Further, research has established that teachers in ED classrooms frequently engage in significantly more negative than positive interactions with students; one study found that during observation, teachers engaged in positive interactions with students only 5% of the time (Sutherland & Wehby, 2001).

Additionally, as previously discussed, special education teacher shortages and the high rate of teacher turnover may contribute to the lack of empirically-based interventions being used in these classrooms. Furthermore, even when teachers are made aware of the types of strategies that have proven successful, many teachers who do attempt to adopt new strategies are unable or unwilling to implement the interventions in the manner in which they were intended (i.e., integrity; Tankersley, Landrum, & Cook, 2004). In general, when teachers are exposed to a new program through professional development, only 5-15% of teachers trained implement the intervention, and those who do typically demonstrate very low levels of integrity (Joyce & Showers, 2002); one study demonstrated that after training, teachers implemented interventions as planned only 4% of the time (Wickstrom, Jones, LaFleur, & Witt, 1998). Additionally, when an intervention is based on principles unfamiliar to the teacher, the level of implementation

and integrity is even lower (Telzrow & Beebe, 2002). Unfortunately, with respect to the class-wide behavior interventions for students with ED, due to the level of difficulty and the number of components involved in the intervention, teachers and staff may be even less likely to implement these interventions with integrity (Telzrow & Beebe, 2002).

**Treatment Integrity**

In experimental research, the measurement of TI has been used to describe the extent to which the independent variable is delivered accurately (e.g., accuracy of implementation; Noell & Gansle, 2013; Gresham, 2009; 2013). Additionally, TI, also referred to as fidelity, reliability, adherence, and implementation (and one of the variables of interest in the current study) has generally been defined as the extent to which an intervention is implemented as intended (Gresham, 2009; 2014). In recent years, the measurement of TI in scientific research has proliferated as researchers have become keenly aware that when empirically-based interventions are used, they may only approximate the outcomes demonstrated in the literature when they are implemented as designed (Sanetti & Kratochwill, 2014). Further, when researchers do not report TI data with their results, it may lead to questions regarding the internal validity of the study (e.g., whether the manipulation of the independent variable was associated with change in the dependent variable; Shadish, Cook, & Campbell, 2002; Noell & Gansle, 2013). Oftentimes, treatment integrity is assumed but not measured; recent literature reviews have found that only 30% of studies published in the Journal of Applied Behavior Analysis included treatment integrity data (McIntyre et al., 2007). However, this figure demonstrates some progress in the field, as it is higher than the 16% of studies reporting

TI data in the Journal of Applied Behavior Analysis reported by Gresham and colleagues in 1993. Regardless, reporting TI data is crucial in research and in practice in order to fully evaluate intervention outcomes.

Over the past 20 years, the idea of treatment integrity has continued to evolve from its original definition to incorporate numerous elements related to acceptability, delivery, and receipt of treatment (Gresham 2009; 2013). Recently, Gresham (2009) suggested that there are a variety of factors relating to whether or not an intervention will be delivered with fidelity, including: (a) ease of implementation, (b) time required to implement, (c) teacher perception of the intervention, and (d) compatibility with the context in which it will be implemented (referring to "resources available, teacher experiences, teacher treatment philosophy, and the instructional environment"; Landrum et al., 2003, p. 152). Additionally, others have proposed that there are, indeed, more factors that will potentially impact TI. Grouped into three broad domains, these are: (a) treatment delivery, (b) receipt of treatment, and (c) treatment enactment (Noell & Gansle, 2013). As a result of these different types of approaches to conceptualizing TI, it has been difficult for researchers to come to a consensus regarding an operational definition of this construct (Noell & Gansle, 2013).

As previously discussed, teachers are more likely to implement interventions that are philosophically in line with their current beliefs about behavior change (Telzrow & Beebe, 2002). In other words, as Long and Maynard (2014) surmised, "if implementation is to occur, a practitioner must have a favorable attitude about the intervention, a belief that other relevant practitioners encourage its use, and a perception that carrying out the

intervention is feasible" (p. 63). If a teacher or staff member does not present with a belief system congruent with the delivery of an intervention, it may subsequently impact intervention fidelity. When this occurs, it requires that the consultant and/or trainer working within the system use various techniques to attempt to improve treatment adherence (Long & Maynard, 2014).

**Measurement of Treatment Integrity**. Although researchers have begun to elaborate various issues related to improving treatment integrity in applied settings, some questions remain. One such unanswered question is with respect to the level of TI necessary to replicate outcomes from research (Gresham, 2013). Previously, researchers had suggested, somewhat superfluously, that 80% was an acceptable level of TI for the successful replication of findings (Mortenson & Witt, 1998; Sanetti & Kratochwill, 2009; Gresham, 2013). However, some have begun to question this number, asking instead about the components of treatment measured, the way in which these components are measured, and the psychometric properties of the measurement tools (Gresham, 2013). Additionally, Gresham noted that one important consideration for the measurement of TI was how much weight each treatment component received. He goes on to explain measurement of all of the components of TI may not be necessary; however, researchers have not yet been able to determine which components of any given program educators need to implement at high levels of integrity to achieve success.

## Professional Development Models

One potential reason for the generally-low levels of treatment integrity observed in prior research is how poorly teachers and staff are trained to implement some

programs. Indeed, despite the lack of empirical support, most districts attempt to provide training for educators utilizing the single-event workshop model (Brock & Carter, 2015; Joyce & Showers, 2002). Typically, these are delivered in a lecture/PowerPoint style without support or programming for the generalization of new skills back to the classroom environment (Joyce & Showers, 2002; Scheeler, 2007). Recently, researchers have been investigating different techniques to facilitate the generalization of skills from the training environment to the classroom.

**Performance Feedback.** Performance feedback (PF) has been used in the field of organizational psychology since the 1970's to change adult behavior across a variety of settings (Alvero, Bucklin, & Austin, 2001). Historically, the term has been used to describe an intervention in which a consultant gives feedback to a staff member or a group of staff in an attempt to improve performance (Alvero et al., 2001). In general, PF is conceptualized as a process in which a consultant monitors specific behaviors related to the treatment integrity of an intervention and then uses data from the monitoring process to provide the consultee with information regarding his or her level of fidelity to the treatment; typically, the consultant provides this feedback in a timely manner and through a graphic representation of the percentage of TI observed. However, Alvero and colleagues (2001) determined researchers have used up to eight different mechanisms to provide feedback to participants, including: graphs, verbal, written, verbal feedback and graphs, verbal and written feedback, verbal and written feedback and graphs, written feedback and graphs, and verbal and mechanical feedback. They also analyzed feedback frequency (how often feedback was delivered), participants (people whose performance

was analyzed for feedback), privacy (how/where feedback was delivered), and feedback content.

In their review of the literature, Alvero and colleagues (2001) suggested that although performance feedback has been used to improve treatment integrity across settings, people, and organizations, the exact mechanism by which performance improves with the intervention of performance feedback is unknown. Some researchers have proposed that PF functions as an antecedent intervention, and yet some others have suggested PF functions as a reinforcer or punisher. Nonetheless, regardless of the function it serves, when performance feedback has been added to an intervention, it has typically been successful in increasing teacher and paraprofessional adherence to the intervention components measured (Alvero et al., 2001; DiGenarro et al., 2005; Westover & Martin, 2014).

Two recent studies have helped to elucidate the extent to which performance feedback can be considered an empirically-supported practice: a meta-analysis looking at the effect of performance feedback on teachers' treatment integrity (Solomon, Klein, & Politylo, 2012) and a systematic review and evaluation applying the What Works Clearinghouse (WWC; Kratochwill et al., 2010) standards for single-case design for performance feedback (Fallon, Collier-Meek, Maggin, Sanetti, & Johnson, 2015). The first study by Solomon and colleagues included analysis of 36 single-case studies where the researcher(s) used performance feedback as defined by Noell and colleagues (2005; e.g., they used data-based feedback delivered verbally or verbally and graphically) to improve teachers' intervention implementation. In this study, Solomon and colleagues

found that across studies, effect sizes varied greatly, ranging from negative values to values close to one. Additionally, they reported that there was a medium to large correlation between the delivery of PF and improvements in teacher behavior. In this way, they surmised that the use of performance feedback overall produced greater improvement in teachers' treatment integrity (e.g., a medium effect) than would be expected if left to chance alone. Further, in their moderator analysis, they found that PF provided to general education teachers typically produced greater increases in TI than PF provided to special education teachers (Solomon et al., 2012).

The Solomon and colleagues (2012) study also included an analysis of student outcomes concurrent with the improvement of teachers' treatment integrity. They reported that 16 of the 36 studies used for the meta-analysis included outcomes related to student behavior (e.g., disruptive behavior, academic behavior, on-task behavior). Overall, the effect sizes for improvement in student behavior ranged from low to high values, with an overall weighted medium effect size. In other words, the improvement in teachers' treatment integrity subsequently produced a "positive and small effect" on student behavior. It is important to note that Solomon and colleagues identified that PF produced a more significant relative change in teacher behavior than student behavior. They indicated that they were not surprised with this outcome given that PF as an intervention is directly related to the fidelity with which the intervention is delivered, and not with how effective the intervention is itself.

Similarly, Fallon and colleagues (2015) recently conducted a systematic review of performance feedback studies in an effort to evaluate whether or not the literature

supports the use of PF as an evidence-based practice. In contrast to Solomon and colleagues (2012), this study assessed the effectiveness of PF by analyzing outcomes of previous studies against the WWC standards for single-case design (Kratochwill et al., 2010). Overall, they found 29 studies, for a total of 102 cases with demonstration of effectiveness that either met WWC standards or met WWC standards with reservations (Fallon, Collier-Meek, Maggin, Sanetti, & Johnson, 2015). Although the evidence standards for single case design will be discussed in more detail in the methods section, overall, Fallon and colleagues determined that due to the overwhelming effectiveness of PF as documented across various studies, PF should be considered an evidence-based practice.

Nonetheless, it is interesting to note here how the Fallon and colleagues (2015) study echoed some of the concerns about PF noted in the Alvero and colleagues study (2001). In both studies, the authors discussed that the mechanism by which PF helped to increase teachers' treatment fidelity was difficult to discern. Because of the variety of ways researchers have applied PF, there is currently no consensus on when feedback should be given (e.g., before an observation, after an observation, a day later), how it should be delivered (e.g., verbally, graphically, written), who should provide PF (e.g., university researchers, administrators, school personnel), and whether or not supplementary activities during PF meetings (e.g., goal setting) could help improve treatment integrity. Finally, and perhaps most importantly, Fallon and colleagues discussed how lack of fidelity assessment tools "may hinder the widespread adoption of this practice [PF] in schools" (p. 241).

**Purpose of the Study**

Landrum and colleagues (2003) argued whether or not the current state of special education services for students with ED is, in fact, special. Indeed, when poorly-trained teachers who lack training and expertise in behavioral interventions are placed in classrooms with students who manifest the most aggressive and violent behaviors, the outcomes are less than optimal. In addition, many researchers have theorized that students with ED do not just require early intervention; rather, they more often need support *throughout* their school careers (Landrum et al., 2003). Additionally, these interventions need to be implemented with frequency and fidelity to ensure effective outcomes for students (Landrum et al., 2003). Finally, in order to be able to implement these interventions with integrity, researchers have surmised that school staff must have access to regular coaching support (Cook et al., 2003). Without this support, teachers will have a difficult time maintaining program changes and fidelity to empirically-based practices (Cook et al., 2003).

Given the dearth of research in performance feedback and empirically-based practices for students with ED, this study was designed to provide a comprehensive analysis of the effect of performance feedback for a class-wide behavioral intervention. As previously discussed, the majority of research in the application of group contingencies with ED students has been conducted in elementary settings; this study is unique in that the recipients of performance feedback are staff in secondary settings. Uniquely, since the integrity of the level system depends on all of the staff members in

the classroom working together, the delivery of performance feedback is directed at all staff members, not just the teacher or paraprofessionals.

This study also followed the research recommendations to evaluate class-wide behavioral interventions outlined by Witt, VanDerHeyden, and Gilbertson (2004). Witt and colleagues recommended that not only should future research be conducted in this specific area, but also that it focus on a three step procedure where (a) preparation of the study consists of the identification of classroom management strategies, training of staff, and the training outcomes are assessed; (b) treatment fidelity of the intervention is monitored, and (c) evaluation of outcomes and student behavior are monitored throughout. As such, this study served to answer the following questions:

1) Is there a functional relationship between the implementation of performance feedback and teachers' treatment integrity to a class-wide behavioral intervention in a secondary school setting?
2) Is there a functional relationship between an increase in treatment integrity of a class-wide behavioral intervention and an increase in student academically-engaged time in a secondary school setting?
3) Is there a functional relationship between an increase in treatment integrity of a class-wide behavioral intervention and a decrease in disruptive student behaviors in a secondary school setting?
4) To what extent are the results of the intervention socially valid?

   a) to what extent do teachers perceive consultation efforts as acceptable?

b) to what extent do instructional assistants perceive consultation efforts as acceptable?

c) to what extent is the degree of behavior change socially significant?

# Chapter 2: Method

## Participants and Setting

Before beginning the project, written consent was gained from teachers and site administrators, and permission to conduct the study was obtained through an application to the district's research unit. Additionally, the researcher gained approval through the university's Institutional Review Board.

The district where this study took place is a large, urban school district in Southern California. Approximately half of students in the district qualify for free and/or reduced cost lunch, and the current truancy rate for the district is 34%. Additionally, during the 2012-2013 school year, the most recent year for which data were available, approximately 400 students in grades 7-12 dropped out of school. Finally, approximately 150 students in the district are eligible for special education under Emotional Disturbance.

This district maintains six classrooms for ED students in middle and high school. For this study, out of the six classrooms, only four were eligible for participation in the study due to issues regarding staffing and administrator concerns. As a result, the sample of classrooms is one of convenience. However, the period in which each classroom was observed for treatment fidelity, student academic engagement, and student disruptive behavior was chosen at random by the researcher. For each classroom, the number of each of the periods in which students were present (e.g., instructional periods) was placed in a box, and the researcher chose one number per classroom. That number indicated the period of day the researcher would visit the classroom to take data. For classroom one,

the period was second and the subject was History. For classroom two, the period was fifth and the subject was English language arts. For classroom three, the period was third and the subject was math. For classroom four, the period was first, and the subject was math.

Each of the four classrooms for students with ED was staffed with two paraprofessionals and one teacher. Teachers ranged in age from early 40's to early 60's, and paraprofessionals ranged from early 20's to early 60's. Additionally, all of the teachers had at least five years of teaching experience; although one teacher just began teaching students with ED two years prior to the beginning of this study. Further, the paraprofessionals in this study ranged in experience from zero years prior experience to 25 years of experience in classrooms with students with ED. However, the average amount of experience for paraprofessionals in this study was 12 years.

With respect to observations, in general, not every student was always present during observation periods due to periodic absences or students leaving the room for services related to their Individualized Education Plan (IEP). Additionally, all students had a special education eligibility of emotional disturbance. The specific demographics of each classroom, along with the staff of each classroom were as follows:

**Classrooms 1 and 3.** Classrooms 1 and 3 were self-contained classrooms for 7th and 8th grade students with ED on the same general education middle school campus. Both classes were staffed with one teacher and two paraprofessionals. The school supported approximately 1,000 students, and the demographic makeup of the school was approximately 5% Black/African American, 1% Native American, 1% Asian, 75%

Latino, and 15% White. Additionally, the majority of students were categorized as socioeconomically disadvantaged, approximately one-third were English language learners, and approximately 10% of students on the campus were eligible for special education. Classrooms 1 and 3 were the two classrooms specifically for students with ED on the school campus, and most of the 16 students with ED rotated between the two classrooms throughout the day.

***Classroom 1***. Depending on the class period, classroom 1 had between two and eleven students. During second period, there were seven students enrolled in the class; however, not every student was always present during the observation period due to periodic absences or students leaving the room for services related to their Individualized Education Plan (IEP). All students had a special education eligibility of emotional disturbance.

*Teacher 1***.** Teacher 1 was an African-American female teacher in her 50s with 11 years of teaching experience. She first received her multiple subject teaching credential and taught students in general education; however, six years ago, she received her special education credential and began teaching in the ED classroom. Teacher 1 had two master's degrees: one in special education and one in cross-cultural education. She did not take any classes specific to behavior management in either of her credentialing programs.

***Classroom 3.*** Similar to Classroom 1, depending on the class period, Classroom 3 had between two and eleven students. During third period, there were five students enrolled in the class.

*Teacher 3***.** Teacher 3 was a Caucasian female in her late 50s. She had been teaching for 16 years, the last five of which were in the ED classroom. Her initial teaching credential was for general education, but five years ago, she obtained her master's in education and a mild to moderate special education teaching credential. She reported taking one to two classes specifically in behavior management as a part of her special education teaching credentialing program.

**Classrooms 2 and 4.** Classrooms 2 and 4 were self-contained classrooms for 9th through 12th grade students with ED on a general education high school campus. Both classes were staffed with one teacher and two paraprofessionals. Approximately 2,000 students were enrolled this high school. Overall, the demographic makeup of the school was similar to the middle school, with approximately 5% Black/African American, 1% Native American, 2% Asian, 66% Latino, and 22% White. Additionally, 71% of students were categorized as socioeconomically disadvantaged, 11% were English language learners, and 11% of students on the campus were eligible for special education. Classrooms 2 and 4 were the classrooms specifically for students with ED on the school campus, and most of the 23 students with ED rotated between the two classrooms throughout the day.

***Classroom 2***. Depending on the class period, classroom 2 had between two and fifteen students. During fifth period, there were eleven students enrolled in the class; however, not every student was always present during the observation period due to periodic absences or students leaving the room for services related to their Individualized

Education Plan (IEP). All students had a special education eligibility of emotional disturbance.

*Teacher 2.* Teacher 2 was an African-American male in his mid-40s. He had been teaching in a special education setting for eight years; however, this was his first year as a teacher of an ED class. He reported taking one or two classes in behavior management during his credentialing program.

***Classroom 4.*** Depending on the class period, classroom 4 had between two and fifteen students. During fifth period, there were ten students enrolled in the class; however, not every student was always present during the observation period due to periodic absences or students leaving the room for services related to their Individualized Education Plan (IEP). All students had a special education eligibility of emotional disturbance.

*Teacher 4.* Teacher 4 was a Caucasian male in his late 50s. He had been teaching for 28 years, and had experience teaching both general and special education. He held a mild/moderate teaching credential, a multiple-subjects teaching credential, and an administrative credential. He had a master's degree and had 18 years of experience teaching ED students. He reported having taken zero classes in behavior management in his credentialing program, but estimated he had about 20 days worth of training in behavior management from the district over the past 28 years.

**Procedure**

As a part of a larger, district-wide initiative, all teachers and staff were provided with two-day training (12 hours total training time) before the school year began on

behavioral principles as a part of a district-wide effort to improve student behavior in the classroom. This training was conducted by two faculty members in the School Psychology department of a large public university in the Southwestern U.S. who specialized in the application of behavioral principles in schools. During the first day of training, a PowerPoint presentation was given during which the presenters introduced principles of applied behavior analysis, and covered topics including operant conditioning, the three term behavior contingency (e.g., antecedent-behavior-consequence), differential reinforcement, and matching law. During the second day of training, the researchers placed an emphasis on the application of these principles to a level system. The teachers and staff were told they were to use the second half of the second day of training to determine the seven components of a point/level system as recommended by Cancio and Johnson (2007). These seven components were: identification of the target behaviors (e.g., behavioral expectations), determination of the point value for behaviors, determination if students earned points for work completion or if there was an opportunity to earn bonus points, development of the continuum of levels (backup reinforcers) students could earn on a daily basis based on the number of points earned per day (see Table 1 for an example of level privileges developed in training), determination of whether or not the classrooms would employ another type of backup reinforcer (e.g., fun Friday), development of a procedure to communicate points/levels with parents, and development of a procedure to monitor student progress.

## Design

This study used a concurrent multiple baseline across classrooms single-case design. In this design, after achieving a stable baseline, the independent variable (e.g., performance feedback) was introduced to one of the classrooms. Subsequently, the IV was introduced to the other three classrooms in a staggered fashion after demonstration of a stable baseline for each classroom.

**Multiple Baseline Design**. The multiple baseline design is the most commonly used single case design in experimental research due to its flexibility and ease of use (Cooper et al., 2007). Additionally, like all single case designs, the multiple baseline design is an experimental design that utilizes within and between subjects comparison to control for threats to internal validity (Horner, Carr, Halle, McGee, Odom, & Wolery, 2005). Further, external validity is accounted for by replication of the effect across subjects and time (Horner et al., 2005). It is important to note that Horner and colleagues (2005) recommend the use of single case design methodology in educational settings for various reasons, including that it focuses on the individual, it allows for detailed analysis of responders and non-responders, and it is cost effective.

A multiple baseline design was chosen for this study because it allowed for an empirical analysis of the intervention across multiple classrooms. Additionally, the nature of the design did not require a return to baseline or removal of an intervention once implemented. This is a unique advantage of the multiple baseline design, and is oftentimes preferred in situations in which removal of an effective intervention could result in the return of undesirable behavior (Cooper et al., 2007).

In single case research, there are inherent threats to internal validity that must be accounted for in order to made accurate conclusions about the functional relationship between the independent and dependent variables (Kratochwill et al., 2010). Possible threats include: ambiguous temporal precedence, selection, history, maturation, regression toward the mean, attrition, testing, instrumentation, and additive and interactive effects (Kratochwill et al., 2010). Multiple baseline designs help to control for the threat of ambiguous temporal precedence by staggering the onset of the independent variable across various points in time (four manipulations in the case of this study; Horner & Odom, 2015). Although the sample of teachers and students was a convenience sample, issues regarding selection, history, and maturation were partially addressed by randomizing the periods of observation and minimizing the length of study.

For this multiple baseline design, there were two phases across four classrooms: (a) baseline/no-feedback phase and (b) treatment/performance feedback. Because this was a multiple baseline design across settings (e.g., classrooms), the dependent variables were measured in all classrooms across baseline and intervention phases. Three dependent variables were measured: staff treatment integrity, student academic engagement, and student disruptive behavior (see dependent variable section for more information).

In addition, this study followed the recommendation standards for multiple baseline designs from the What Works Clearinghouse (WWC; Kratochwill et al., 2010). Based on their recommendations, to meet standards with reservations, at least three observations are required for each phase (baseline and treatment) to accurately gauge the

trajectory of the data. When using a multiple baseline design, Kratochwill and colleagues also recommend including a minimum of six phases; since four classrooms will be chosen to participate in this study, in the absence of attrition, there will be eight phases. In other words, each baseline phase for each classroom is a separate phase ($n$ = 4), and each treatment phase for each classroom is a separate phase ($n$ = 4).

**Baseline**. During the initial stages of the intervention when the baseline/no-feedback phase was in effect, the teachers were asked to implement the intervention as trained. The consultant was in each classroom for approximately one to two hours each week observing students and teachers, helping to create materials for the program as teachers requested, and helping to tally students' daily and weekly level points. Additionally, the consultant met with all of the teachers and staff one time per week during their weekly professional development time. These weekly meetings were arranged by site administrators to review policies, program issues, and address staff's questions. The consultant used approximately 20-30 minutes out of this hour-long meeting to answer general questions regarding the intervention, and review general behavioral principles, but did not provide specific feedback on implementation efforts. Examples of questions asked by staff during this time were "Can students earn points towards their levels outside of the classroom?", "Should students continue to accrue points when they are out of the classroom receiving IEP-related services?", and "How often can we change the rewards offered for each level?" Additionally, during this phase, staff completed an Excel spreadsheet detailing the number of points students earned each

day. The consultant used this data to compile weekly summary data to review with the teams on the percentage of points earned by student by week and by month.

Throughout the baseline condition, the consultant also worked with teachers and staff to help assess the function of certain individual students' behavior for purposes of designing and implementing individual behavioral interventions. Additionally, for certain students who already had individualized behavior plans, the consultant worked with the school teams to determine if they were helping to improve student behavior.

**Intervention/Performance Feedback Phase**. The primary researcher determined the order in which classrooms began receiving intervention based on a visual analysis of level and trend of the baseline TI data. With a multiple baseline design, intervention should begin (e.g., the introduction of the independent variable) after baseline data are stable or trending in the opposite direction from that which is expected during intervention (Cooper et al., 2007; Kratochwill et al., 2010). As a result, after at least three data points were collected the data was examined for stability and/or trend. After the third data point was collected in Classroom 1, the baseline phase was identified as producing a stable rate of responding via a visual analysis (i.e., the data path was horizontal to the x-axis). Intervention began in Classroom 2 during week 7 after the last three data points collected all indicated the same level of TI (e.g., a stable rate of responding). For Classroom 3, intervention began during week 8 after TI data for the last two data points were relatively stable at 37% and 40% TI, respectively. Finally, for Classroom 4, intervention began during week 9 after the last four TI data points consistently varied between 15 and 30%.

## Independent Variable

**Performance Feedback**. During the PF phase for each classroom, the consultant met with the staff in a group one time per week at the staffs' convenience. For classrooms 1 and 3, meetings occurred after school on their early release day (Wednesdays), and for classrooms 2 and 4, meetings occurred during the students' lunch period on Thursdays. Treatment integrity data was collected on Mondays and Tuesdays for all of the classes using the Treatment Integrity Observation System (TIOS), a researcher-developed measure, and the data from this observation was shared with the team during their weekly meeting. The weekly meetings were held regularly with the classroom staff, teachers, and site administrator(s) to discuss program concerns and updates in one of the ED classrooms on the school campus. The meeting day differed according to the site; the middle school staff met on Wednesdays, and the high school staff met on Thursdays. At both sites, the meetings were scheduled for an hour, but typically only lasted about 30 minutes. The consultant sat in on those meetings with the administrator. When the administrator finished his or her part of the meeting, the consultant then left the classroom where the meeting was held to conduct PF with one group of staff in their classroom. After conducting PF with the staff of one classroom, the consultant returned to the original meeting location to conduct the PF session with the other classroom.

In each classroom, there was a tabulated binder with (a) information about the level system presented in the original training (i.e., the PowerPoint presentation, handouts, and notes), (b) empty data sheets staff used to capture students' daily points with instructions, and (c) information on the levels (e.g., how many points students

needed to earn for each level, how levels were calculated); when the PF condition began, the consultant added (d) information on how integrity was collected and the consultant's role in gathering integrity. Included in Tab D was the TIOS along with criteria for obtaining full credit for each section (i.e., definitions of what the consultant was looking for in each section). In this way, staff could review this information regularly if needed.

With respect to what was presented to staff in the meetings, the results from each observation were graphed using Excel and displayed as a percentage of TI for each component measured. The graphs were printed out for the staff to review during the meeting; there was one master copy that the consultant used to show the group the breakdown of points, and each staff member was provided with one graph each on which they were able to take notes or write down information. During the meeting, the consultant verbally explained the different components and gave staff specific feedback as to why scores were high or low in each area. For the components in which staff scored less than 100% TI, the consultant gave specific information about what the team could do to improve TI (e.g., in the permanent product review section, the consultant might say, "levels need to be posted with descriptors of what students have access to at that level"). Also, during each session, staff were able to view their overall treatment integrity over time and were told whether it was increasing, decreasing, or staying the same. Finally, the consultant allowed time for questions. Before the consultant left the meeting, the consultant placed one master copy of the graphs and materials reviewed in the meeting in the binder under Tab D that remained in the classroom for staff to review as necessary. Overall, meetings lasted from 9-15 minutes.

Procedural integrity for meetings was gathered during all PF sessions. The consultant completed an eight-item checklist during each session as each component of the PF session was completed. The items on the checklist included greeting members of the team and providing an overview of the meeting, providing the most recent TI graph for all members to review, verbally providing information regarding percentage of TI achieved while referencing the information on the graph, providing staff members with graphs measuring their TI progress over time, verbally providing information on staffs' average percentage of integrity over time, setting goals for the next meeting, verbally answering all questions from staff, and maintaining a time limit of 15 minutes.

Overall, the consultant's ratings on the delivery of PF ranged from 83 to 100%, for an average of 92% integrity. Additionally, 24% of PF sessions included an outside observer who collected PF data with the same procedural integrity checklist. Inter-observer agreement (IOA) was calculated as percent agreement (number of agreements divided by total number of agreements and disagreements), or the degree to which two observers agreed on whether or not the consultant engaged in a specific behavior. For this project, IOA was 100% across all sessions. See Figure 7 for the procedural integrity checklist.

**Dependent Variables**

The dependent variables were treatment integrity, measured as the percentage of total points staff earned per observation period, student engagement, and student disruptive behavior. Treatment integrity was the primary dependent variable and student engagement and disruptive behavior were chosen due their relationship to student success

and potential sensitivity to change (McKissick, Hawkins, Lentz, Hailley, & McGuire, 2010)

**Treatment Integrity**. Treatment integrity was the primary dependent variable of interest in this study. As previously discussed, treatment integrity is the extent to which the intervention is delivered as intended. Since the components of this class-wide intervention were designed specifically to increase student engagement and decrease student disruptive behavior, in order to evaluate student outcomes effectively, high fidelity of implementation was of utmost importance. In other words, not only was treatment integrity the main dependent variable targeted for change through PF, it was also a key variable in determining the extent to which the intervention may be associated with improvement in student behaviors. For information regarding the measurement of treatment integrity, see the explanation of the Treatment Integrity Observation School (TIOS) in the measures section.

**Student engagement.** Due to the different types of activities involved in academic learning, the construct of student engagement has been difficult for researchers to define and measure (Jimerson, Campos, & Greif, 2003). However, defining engagement in the context of a behavioral framework has given researchers an opportunity to directly and objectively measure this behavior. In this context, academic engagement can be measured through students' overt behavior in participatory activities such as hand raising, orienting towards the teacher during a lesson, and complying with academic demands (e.g., writing in a notebook when asked to do so, reading aloud; Downer, Rimm-Kaufman, & Pianta, 2007). In other words, by looking for and directly

observing these types of behaviors, researchers have developed tools to measure student behaviors that enable academic learning (Greenwood, Horton, & Utley, 2002).

Using structural equation modeling, Greenwood (1996) determined that when student engagement was defined behaviorally, a causal relationship was supported between student engagement and subsequent academic success. Further, not only was engagement related to students' classroom performance, it was also a mediator between instruction and overall student outcomes (Greenwood, 1996). In other words, a model that included student academic engagement provided a more complete picture when looking at the relationship between instruction and school outcomes. Although this mediating relationship is important to consider for all students, it is especially important for students at risk for failure due to emotional and/or behavioral difficulties and those from economically disadvantaged backgrounds (Downer et al., 2007; Howse, Lange, Farran, & Boyles, 2003; Vile Junod, DuPaul, Jitendra, Volpe, & Cleary, 2006). For example, using a behavioral framework to measure student engagement, Vile Junod and colleagues (2005) determined that students with Attention Deficit Hyperactivity Disorder (ADHD) exhibited significantly lower rates of academic engagement and higher rates of off-task behaviors than their peers during reading and math instruction.

Finally, it is important to note that when engagement is defined behaviorally, not only is it observable and measureable, it is also readily alterable (Greenwood, 1996). Both of these qualities make academic engagement a uniquely appropriate outcome variable to assess in studies seeking to improve outcomes for students with ED. Few studies have utilized the measurement of class-wide student academic engagement as an

outcome measure for contingency management, but of those that have, group contingencies have produced some positive effects on academic engaged time (McKissick et al., 2010; Riley-Tillman, Methe, & Weegar, 2009).

**Disruptive behavior.** Student disruptive behavior (e.g., talking out of turn, throwing items, engaging with other students  of academic tasks) is often times incompatible with academic engagement (Downer et al., 2007; Greenwood et al., 2002). Additionally, in ED classrooms, disruptive behavior is one of the most salient barriers to student learning (Landrum et al., 2003). Further, responding to and reactively managing disruptive student behaviors is reported as one of the more stressful activities for teachers (McKissick et al., 2010), and the amount of disruptive behavior a student engages may affect how staff in the classroom interact with him or her (Sutherland & Oswald, 2005). In addition, in classrooms where there are a number of students with disruptive behaviors, some have hypothesized that teachers are prevented from delivering high-quality instruction (Wehby et al., 2003). Therefore, targeting this variable for change could have large implications on how teachers perceive the intervention's effectiveness.

**Measures**

Treatment integrity data were collected through the observational and permanent product measures detailed in the Treatment Integrity Observation System (TIOS), a system of measuring TI created by the author; Table 2 lists the components of TI measured by the TIOS. Additionally, student academic engagement and student disruptive behavior data were collected through the use of the Classroom Behavior Observation Tool (CBOT), a system of collecting class-wide data developed by the

author based on adaptations of both the Behavior Observation System of Students (BOSS; Shapiro, 2004) and the Planned Activity Check (PLACHECK; Cataldo & Risley, 1973; Cooper et al., 2007).

**TIOS Development**. Items for the TIOS were chosen based on the concepts presented in the two-day training and manualized components of the point/level class-wide intervention. The development of the treatment integrity tool was modeled after the tool developed by Jeffrey and colleagues (2007) in their examination of class-wide treatment integrity. In their study, they categorized core classroom management studies into four domains: classroom ecology, materials, teaching expectations, and instructional management. They identified the various strategies associated with each domain (e.g., teacher praise, using point card) and how data would be collected for each component (e.g., direct observation, permanent product). They then used the tool they developed to collect data and provide performance feedback to four elementary and five middle school special education teachers for students with ED.

Initially, the author systematically reviewed the training and manual materials presented to the staff at the two trainings. As previously mentioned, participants left the training with tabulated binders, including (a) information about the level system presented in the original training (i.e., the PowerPoint presentation, handouts, and notes), (b) empty data sheets staff used to capture students' daily points with instructions, and (c) information on the levels (e.g., how many points students needed to earn for each level, how levels were calculated). Then, similar to the Jeffrey and colleagues (2007) study, the researcher outlined three major components of the system: (a) staff use of behavioral

strategies, (b) materials, and (c) teaching and reinforcing of behavioral expectations. In order to capture the various components of the intervention, the TIOS was broken into two distinct components: permanent product review and an observational system. The permanent product components reviewed were unique to the class-wide intervention developed by the trainers, and included posting of class rules, levels posted with corresponding student names, and use of the point system. For the permanent product observation, for both the posting of the class rules and the levels posted with corresponding student names, the author would look around the room and indicate whether or not each component was present. The data sheets were collected each time the author entered the classroom for an observation period, and the datasheets were reviewed to ensure that all data was entered for the previous week for each student in the classroom. Table 1 describes each domain, the tasks involved, and the method of measurement.

The direct observation components of the TIOS included (a) a system to indicate whether or not certain components were being utilized by staff (i.e., level privileges such as board games or mechanical pencils were evident in the class environment *and* students were accessing them during the observation time period), and (b) a direct 15 minute observation period that required the observer to indicate the presence or absence of certain components of the system via 30 second partial interval recording. During each 30 second period, the presence or absence of an opportunity in which a student was engaged in an appropriate or inappropriate behavior was recorded. Appropriate student behaviors were identified as one or more students following staff directions, remaining in their seat,

engaged in an academic task, or following a classroom rule. If one or more students were engaged in one of these behaviors at any point within the 30 second period, this was indicated, and the observer would identify if any staff member (a) provided verbal praise to a student, (b) provided praise that was behavior specific (e.g., "good job following directions and taking out your book"), and (c) marked that the student earned a point for engaging in appropriate behavior. Conversely, if the observer identified that one or more students in the classroom were breaking a classroom rule, this was indicated, and the opportunity for staff to engage in the use of correction procedures was then present. In other words, when a students was not following one or more of the classroom rules, it was an opportunity for the staff to begin the progression of behavior correction procedures. The five components of the correction procedure were (a) staff giving a verbal directive (i.e., prompt) for the student to engage in the appropriate behavior and waiting five seconds, (b) staff giving a verbal directive again, (c) staff telling the student they were on a time out from points for a specified period of time, (d) staff making a corresponding mark on the data sheet indicating the time out from points period, and (e) while the student was on a time out from points, staff continued to praise student. If, during the progression of staff remediation for student inappropriate behavior, the student began to again engage in appropriate behavior, the staff member was no longer required to move through the response hierarchy. In other words, if staff engaged in the first response where they gave the student the verbal directive to engage in appropriate behavior, and the student began to behave appropriately, the opportunity for staff to use the behavior correction procedure would end.

Construct validity for the TIOS was gathered via feedback from an expert review panel, a subsequent revision process, a second review, and then a final revision process. The expert panel consisted of four professors with Board Certifications in Behavior Analysis, Doctoral level (BCBA-D), eight graduate students in a master's program in Applied Behavior Analysis, twelve graduate students in a doctoral program in school psychology, the program director of a doctoral program in school psychology, and ten masters-level school psychologists. For the initial stage of gathering feedback, the researcher spent an hour and a half presenting the proposed research project to the four professors and eight graduate students in the BCBA program. During this presentation, the professors and students could ask clarifying questions as needed. The researcher emphasized the importance of understanding the project in order to gain appropriate and relevant feedback regarding the proposed measures. After the presentation, the first versions of the researcher-created measures were passed around to the students and faculty to review. The participants spent approximately 20 minutes reviewing the materials and discussing them with each other while marking their comments and ideas for revision on the forms. The researcher remained present during this time period to answer any of the participants' questions. After the review period, the participants all handed the measures with their comments back to the researcher. Additionally, one of the professors emailed further comments and suggestions to the researcher later that week.

With respect to the feedback for the measures from the twelve graduate students in a doctoral program in school psychology and the ten masters-level school psychologists, the information provided to these participants was much more brief in

nature than with the other participants. For these participants, the researcher met with the groups for 10 minutes to outline the project, and asked the two groups to evaluate the forms on the feasibility of their use in a school setting. All participants reviewed the measures on their own and then emailed their comments and suggestions to the researcher within one week.

After receiving all of the comments and feedback from the participant groups, the researcher made changes to the original version of the measures. For the TIOS, the researcher removed an interview section where the objective was to ask a student in the class to name the rules of the classroom. Specific feedback on this component of the measure included that this was not measuring the extent to which classroom management procedures were in place; rather, it measured the understanding of one of the students in the classroom. The reviewer that identified this issue felt that improvements in this area were only indicative of student learning, and did not have to do with whether or not the classroom management system was being implemented. Additionally, as initially conceptualized, the initial direct observation portion of the TIOS consisted of three five-minute partial interval recording opportunities. Many reviewers identified that five minutes was too long, and that the researcher would be unable to accurately capture staff behavior as a result. The suggestions for remedying this included shortening the interval to 30 seconds and only observing staff behavior if there was an opportunity to respond (i.e., a student engaged in inappropriate or appropriate behavior during the specified time period).

Once these changes were made to the measure, the researcher sent out the revised form to all of the participants to solicit their feedback on the improved measure. The participants were asked to review the form and give any feedback to the researcher only if they identified something that could be further improved. Three participants responded with further comments around helping to streamline the praise ratio scoring process and making the forms more user-friendly (i.e., keeping the permanent product observation and direct observation part one on the first page of the measure).

After the second revision, the measure was piloted across all ED classrooms in the district, and interrater reliability was collected. The researcher identified three other graduate students in a school psychology doctoral program who were willing to help with piloting the data and trained them on the measure. These three students were in their first year of the doctoral program. The training included an in-depth explanation of the project, the type of data to be collected, and the rationale for why those specific components of the classroom management system were chosen. The training lasted approximately one and a half hours, after which the students were encouraged to ask clarifying questions. Next, the three students watched as the researcher used the measure to collect data in one of the classrooms while following along with their own TIOS forms. As they watched, they wrote questions down on their respective forms which the researcher was able to answer after the observation period. After the researcher answered their questions, one-by-one each student accompanied the researcher to an ED classroom and scored the TIOS alongside the researcher twice for two different class periods. When each student had scored the TIOS four times alongside the researcher, the researcher and

each student visited each ED classroom in the district to complete the pilot observations. When conducting a pilot observation, the researcher and the other student stood in the back of each classroom on opposite sides of the room. Interrater agreement was then calculated for all observations. Pilot interrater reliability for the measure was 93% across raters and classrooms.

During the PF phase of the study, the researcher observed and collected integrity data using the TIOS. With respect to interrater reliability, the same three graduate students were enlisted to collect these data throughout the duration of the study. Before conducting observations for interrater reliability, the researcher reviewed the forms with the graduate students and answered any of the student's questions. Additionally, the researcher gave the students an opportunity to practice collecting data with the form if they felt it necessary; however, none of the students chose to do so.

**CBOT Development.** The development process for the CBOT followed a similar pattern as for the TIOS; construct validity for the CBOT was evaluated via feedback from an expert review panel, a subsequent revision process, a second review, and then a final revision process. The measure was initially developed by the first author based on the format of the BOSS (Shapiro, 2004). With the CBOT, both student academic engagement and student disruptive engagement were measured via interval recording. Similar to the BOSS, student academic engagement was measured at the end of each 30 second interval (momentary time sampling), and student disruptive behavior was measured separately, via partial interval time sampling at 15 second intervals. Instead of sampling from one student at a time, however, as the BOSS recommends, all student behavior was observed

at once in a fashion detailed in the PLACHECK (Cataldo & Risley, 1973; Cooper et al., 2007).

The planned activity check was designed to provide an observer with the behavior of a whole group of participants at one time while they engage in defined activities (Barnett, Lentz, Bauer, Macmann, Stollar, & Ehrhardt, 1997). In their description of the procedure, Barnett and colleagues identified that participants' engagement in activities could be determined via an observation procedure whereby the observer first counts the number of students present and then systematically determines the number of participants engaged in the activity during a pre-determined time interval. For each time interval, the number of students engaged is divided by the number of total students to give the percentage of students engaged per interval. The percentage of students engaged per interval is summed and then divided by the number of intervals to give the average percentage of students engaged during the observation period.

In the development of the CBOT, the researcher took deliberate measures to consider both the topographical and functional definitions of engagement and disruptive behavior. Additionally, the definitions for both academic engagement and disruptive behavior were reviewed by the expert panel during the development of the measure. As a result of this process, the researcher created the following definitions:

*Academic engagement* - Student was demonstrating any of the following behaviors: using classroom materials according the lesson, writing, reading, raising hand, orienting towards the teacher during a lecture, talking to the teacher

about assigned material, talking to a peer about the assigned material, sitting in assigned seat/area.

*Disruptive* - Disruptive behavior included any audible noise or movement and/or involved the student physically or verbally engaging with another person in the room that removed them from engaging in the task. The behavior only classified as disruptive if it caused the teacher or other students to look in the student's direction, and/or interrupted the teacher or other students. Examples included making audible noises, calling out, making unauthorized comments, hitting others, throwing objects

Before each observation using the CBOT, similar to the procedure for the PLACHECK, the researcher counted the number of students present. For academic engagement, 80% of the number of students present were calculated. For example, if five students were present, 80% of five was calculated. In this scenario, if four students (80% of the class) were engaged at the end of each interval, the researcher indicated that with a plus sign. However, if less than 80% of students were engaged, the researcher indicated as such with a minus sign. At the end of the observation period, the number of intervals with plus signs was divided by the total number of intervals to determine the percentage of intervals in which at least 80% of students were engaged.

Eighty percent was used as the benchmark for class-wide engagement based on the discussion of engagement rate by Gettinger and Seibert (2002), who suggested that there are specific variables teachers can manipulate in order to maximize academic learning time of students. In classrooms where teachers maximized academic learning

time, students were actively engaged approximately 80% of the time. This minimum level of academic engagement appears to be necessary for maximizing student outcomes. The final CBOT is included as Appendix E.

During the PF phase of the study, the researcher observed and collected data on student academic engagement and disruptive behavior using the CBOT. With respect to interrater reliability, the same three graduate students as were involved in the piloting of the TIOS were enlisted to collect these data throughout the duration of the study. To train the students on the CBOT, the three students watched as the researcher used the measure to collect data in one of the classrooms while following along with their own CBOT form. As they watched, they wrote questions down on their respective forms which the researcher was able to answer after the observation period. After the researcher answered their questions, one-by-one each student accompanied the researcher to an ED classroom and scored the CBOT alongside the researcher twice for two different class periods. When each student had scored the TIOS four times alongside the researcher, the researcher and each student visited each ED classroom in the district to complete practice observations. When conducting practice observations, the researcher and the other student stood in the back of each classroom on opposite sides of the room. Each student was considered proficient in using the measure after three consecutive ratings with 90% or higher interrater reliability with the primary researcher.

**Procedural Fidelity.** Procedural fidelity was also collected for the delivery of performance feedback by the author on the procedural checklist described earlier. Throughout each meeting, the author would monitor the delivery of all of the components

of performance feedback by checking off each item as it was performed. After each session of meeting with the team, the author would add up the number of items performed according to the checklist and divide that number by 8 (the total number of components), and then multiply that number by 100. Overall, performance feedback was delivered with 98% fidelity. Additionally, interobserver agreement was collected during 22% of PF sessions. The same three graduate students who collected interrater agreement data for the CBOT and TIOS also helped to collect interrater agreement data for procedural fidelity. No formal pretraining was necessary as the students were familiar with the system, and the procedural fidelity form was a checklist. However, the researcher gave the students the form one day in advance and encouraged them to ask clarifying questions if necessary. Interobserver agreement for PF delivery was 100%.

The procedural integrity checklist for performance feedback is included as Appendix G.

**Social Validity**

Kazdin (1977) first discussed social validity as it applied to the field of Applied Behavior Analysis (ABA); he determined that social validity should be considered as distinct from empirical significance. He offered that due to the use of single case design in ABA, researchers were criticized for the inability to convert findings into statistically significant results. As a result, Kazdin proposed that through the process of analyzing social validity, the rigor of scientific findings in single case design would be improved. In this way, socially valid outcomes would include three distinct elements: social significance (the extent to which behaviors targeted for change are socially relevant),

social appropriateness (the extent to which the treatment procedures are socially appropriate), and social importance (the extent to which the change in behavior is clinically significant) (Kazdin, 1977). Additionally, overall effectiveness of an intervention would be evaluated through both subjective methods (e.g.,questionnaires) and social comparison methods (Finn & Sladeczek, 2001).

Horner and colleagues (2005) expanded this discussion to include selection of dependent variables that have high social importance, demonstration that the IV can be applied with fidelity by intervention agents, demonstration that the intervention agents charged with delivering the IV report the procedure to be acceptable, feasible, and effective, and demonstration that the effect of the intervention is clinically significant. Based on the previous discussion of the current outcomes for students with ED, it would appear that the selection of the dependent variables of student engagement and student disruptive behavior are socially valid. Additionally, since improvement in TI has been shown to improve student outcomes (Gresham, 2013), the dependent variable of TI is also appropriate.

**Treatment Acceptability.** Staff treatment acceptability, which is one component of social validity, was calculated for both the level intervention and the consultant's role in providing performance feedback. To measure this, the Intervention Rating Profile-15 was used (IRP-15; Martens, Witt, Elliott, & Darveaux, 1985). The IRP-15 assesses teachers' acceptability of an intervention and is a shortened version of the Intervention Rating Profile (IRP; Witt, Martens, & Elliot, 1984). Although the IRP-15 was developed 30 years ago, it remains one of the most widely used measures of treatment acceptability

(Finn & Sladeczeck, 2001). The IRP-15 includes 15 statements (e.g., “I liked the procedures used in this intervention.”) to which participants respond on a six-point Likert scale. The IRP-15 was designed to measure “general acceptability” and in a principal components factor analysis, it yielded one primary factor with item loadings rating from .82 to .95 (Martens et al., 1985). Additionally, in Martens and colleagues (1985) original study on the IRP-15, they found reliability using Cronbach’s alpha was .98. In the current study, Cronbach’s alpha for staff responses for the level intervention was .94, and for the consultant was .90. See Appendix F for the IRP-15.

# Chapter 3: Results

## Interobserver Agreement (IOA)

Based on Kratochwill and colleagues (2012) recommendations, IOA was collected during at least 20% of all dependent variable observation sessions in each phase. Observers were all graduate students in a school psychology program. IOA was calculated using interval-by-interval IOA, where the primary and secondary's observational data for each interval are matched and compared (Cooper et al., 2007). This type of IOA is appropriate when measuring behavior via time-sampling methods (Cooper et al., 2007). Interval-by-interval IOA was used for the time sampling portion of the TIOS, and both variables measured through the CBOT: academic engagement and disruptive behavior. The following formula represents how this IOA was calculated:

$$\frac{Number\ of\ intervals\ agreed}{Number\ of\ intervals\ agreed + Number\ of\ intervals\ disagreed} \text{ x } 100$$

Minimum acceptable values of IOA are at least 80% (Kratochwill, Hitchcock, Horner, Levin, Odom, Rindskopf et al., 2012). IOA for all measures is reported in subsequent sections.

**Treatment Integrity.** In classroom 1, during baseline, IOA was collected during 100% *(n* = 3) of observation sessions, and ranged from 76 to 90%, for an average of 83%. For the treatment phase, IOA was collected for 29% (*n* = 2) of observation sessions, and ranged from 86 to 100%, for an average of 93%.

In classroom 2, IOA was collected during baseline in 50% (*n* = 2) of observation sessions, and ranged from 80% to 90%, for an average IOA of 85%. Additionally, IOA

was collected during intervention in 25% (*n* = 1) of observation sessions, for an IOA of 82%.

In classroom 3, IOA was collected during baseline in 50% (*n* = 3) of observation sessions, and ranged from .84 to .89, for an average of .86. For the treatment phase, IOA was collected in 25% of sessions (*n* = 1), for an IOA of 100%.

Finally, for classroom 4, IOA was collected in 50% of observation sessions during baseline (*n* = 3). Agreement ranged from 84 to 90%, with an average of 88%. During treatment, IOA was collected during 75% (*n* = 3) of observation sessions, and ranged from 78 to 97%. Average IOA for treatment was 90%.

**Academic Engagement/Disruptive Behavior.** Inter-observer agreement for classroom one was calculated for 30% (*n*=3) of overall observations for an IOA of 92%. For classroom two, IOA was calculated for 22% (*n*=2) of all observation sessions at 90%. IOA was calculated for 40% (*n*=4) of observation sessions at 90% in classroom three. Finally, IOA was calculated for 60% (*n*=6) of sessions in classroom four, for an IOA of 89%.

**Analysis**

**Visual Analysis**. Evidence of the functional relationship between the improvement in treatment integrity and student behavior during the intervention phases should be evidenced by the application of baseline logic, including prediction, verification, and replication (Cooper et al., 2007). Cooper and colleagues describe baseline logic as a tool to interpreting the results of the study; in a multiple baseline study, prediction refers to the continuation of the baseline trend without intervention,

verification occurs when one of the graphs evidences a change in the dependent variable congruent with application of the independent variable, and replication occurs when another graph demonstrates the same relationship (see visual analysis section for more information).

Additionally, the visual analysis of data includes the assessment of six outcome-measure data both between and within phases. These include (a) level, (b) trend, (c) variability, (d) immediacy of effect, (e) overlap, and (f) consistency of data patterns across similar phases (Kratochwill et al., 2012).

With respect to level, the average level (horizontal average of lines) should be markedly different between baseline and intervention phases. Additionally, if the trend in baseline is downward, a positive effect would be evidenced by an upward trend for TI and AET during the intervention phase. Conversely, for disruptive behavior, if the trend during baseline is upward, a positive effect would be evidenced by a downward trend the intervention phase. Regardless, there should be little variability (spread of data points) during the intervention phase for all dependent variables, and the effect of the intervention (difference in level and trend) should be apparent soon after the intervention begins. Finally, there should be little overlap between data points between the baseline and intervention phases, and intervention data should be consistent across classrooms for all dependent variables.

Kratochowill and colleagues (2012) relate that if all of the six criteria are met, there is ample evidence to indication that there is a functional relationship between the IV and the DV. Nonetheless, for a study to provide strong evidence of an effect, there must

be “at least three demonstrations of the intervention effect, each occurring at a different point in time, combined with no failures to observe effect” (p. 32). If, however, there is at least one demonstration where no effect is evident, the ability to conclude whether or not a functional relationship exists is jeopardized (Kratochwill et al., 2012; Horner et al., 2005). Tables 3-5 detail changes between baseline and treatment across variables.

**Quantitative Analysis.** Although visual analysis is frequently used in single case designs, there have been concerns regarding its accuracy, especially when the data are variable in the baseline and/or intervention phases (Brossart, Parker, Olsen, Mahadevan, 2006). As a result, Kratochwill and colleagues (2010) suggest including parametric and/or non-parametric effect size (ES) estimates. It is important to note, however, at this time, researchers have not come to an agreement on which types of statistical analyses are most appropriate (Kratochwill et al., 2012). Additionally, researchers currently caution against placing too much weight on parametric estimates for single-case design because they may potentially violate the assumption of independence of errors (Kratochwill et al., 2012). Nonetheless, Parker and Hagan-Burke (2007), relate that effect size estimates, although not perfect, can lead to a more objective picture regarding intervention strength similar to that of group studies. In addition, Kratochwill and colleagues argue the inclusion of ES estimates helps with the credibility of single case research findings, and regardless of the ES used, one could be fairly confident of rank ordering of intervention strength based on the ES coefficient. Finally, analysis of ES can provide a basis for comparison of similar studies (Kratochwill et al., 2012)

There are currently various ways to analyze and report ES in single case research (Parker, Vannest, & Davis, 2011). As a result, multiple effect sizes are typically reported. The first method, a nonparametric ES analysis, although a crude estimate of effect size, can be a helpful first step in determining the extent to which a functional relationship between the IV and the DV exists (Parker et al., 2011). Parker and colleagues (2011) suggest that nonparametric/nonoverlap methods are appealing due to the fact that they are able to help quantify the process achieved through visual analysis. In addition, these methods do not require "parametric assumptions about data distribution or scale type" (Parker et al., 2011, p. 305). Parker and colleagues suggest visually analyzing the data before determining which method to use in consideration of trend; in order to apply one of the nonoverlap methods appropriately, the results may need to include a presence of a strong upward or downward trend (depending on the expected trend for intervention), and the presence of a sharp upward or downward trend (opposite to that of the baseline phase) during the intervention phase.

It is important to note that although various nonparametric ES estimates are available, both percentage of nonoverlapping data (PND) and percentage of the median (PEM) are generally not recommended for use. Kratochwill and colleagues (2012) do not recommend the use of PND due to its undesirable statistical qualities, inability to capture trend, and inability to measure an overall effect. Further, because of the issues related to PND, Parker and colleagues (2015) determined that PAND was created specifically to help remediate the weaknesses of PND: the overemphasis on a single datapoint and lack of a sampling distribution (Parker et al., 2015). Additionally, PEM was not considered

because it can distort results if the data are lacking in central tendency, it generally lacks statistical power, and it is insensitive to data trending upward or downward at the high and low end (Parker et al., 2011).

*PAND/Phi.* In this study, the nonparametric effect size reported is percentage of all nonoverlapping data (PAND; see Figure 1). In addition to the issues previously discussed, the use of PAND is generally supported over other nonparametic ES estimates because it can be converted into a Pearson *Phi* (Φ) coefficient. *Phi* is a parametric ES estimate, which is a regression-based estimator of ES (Maggin et al., 2011). Additionally, unlike the other nonparametric ES estimates, it takes all of the data points from the intervention phase into account (Parker, Hagan-Burke, & Vannest, 2007). Although there are advantages to reporting PAND, there are several limitations that must be considered. First, due to the fact that only the highest or lowest baseline data point is used for comparison to the intervention data, PAND can be insensitive to outliers in the baseline phase. Additionally, for accurate calculation of PAND, 20 data points are suggested (Parker et al., 2007). Finally, Maggin and colleagues (2011) report that all of the nonparametric ES measures are problematic because of their inability to "account for data trends, represent the magnitude of treatment effect, and produce a known sampling distribution" (p. 304).

In order to calculate PAND, the following was calculated: the number of intervention points that overlap with the baseline points divided by the total number of data points and then subtracted from 100%. After PAND was calculated, the results were converted to Pearson's *Phi* (Φ), which is similar to Pearson *R* for a 2 x 2 contingency

table (see Figure 2). Additionally, the effect sizes can be interpreted in standard deviation units, similar to those used within group design studies. An added advantage is that Cohen's (1988) guidelines for interpreting ES may be used to analyze the results (Parker, Hagan-Burke, & Vannest, 2007). With respect to Cohen's (1998) ES estimates (*d*), 0.2 is considered a small effect, 0.5 medium, and 0.8 is a large effect (Parker et al., 2007).

PAND was calculated for all IVs by counting the number of data points in the intervention phase that overlap with any of the data points from the baseline phase. PAND for all variables was calculated for all classrooms compared below. To calculate PAND, similar to the Parker and colleagues (2007) study, a non-overlap line was created and extended vertically through the intervention phase from the highest data point from the baseline phase. Then, the number of data points that overlapped was divided by the total number of data points and then subtracted from 100 (Parker et al., 2007). Since PAND calculations result in a 50-100 scale, where a PAND of 50 is chance level, conversion to a 0-100 scale was achieved by multiplying PAND by 2 and then subtracting 1. Finally, the guidelines provided by Parker and colleagues were used to interpret the PAND for effectiveness of the study as detailed below.

*Phi (Φ)*. *Phi* is calculated using a 2x2 table with the data where, after eliminating the overlapping points between baseline and intervention phases, two ratios are created (Parker et al., 2007). The two ratios are: (a) half of all removed baseline data points divided by the remaining (lower) data points, and (b) the remainder (higher) of intervention data divided by one half of all removed data points. Next, the following equation is applied: $\Phi$ = [a/(a+c)-[b/(b+d)] (Parker et al, 2007). *Phi* was calculated for all

classrooms separately. *Phi* can also be calculated by (PAND * 2)-1 when a completely balanced table is present, in that the percentages of baseline and intervention phases are equal (Parker & Hagan-Burke, 2007). Interpretation of *phi* is through the use of Cohen's (1998) guidelines.

*IRD.* In a recent analysis, Parker and colleagues (2015) suggested that Improvement Rate Difference (IRD) is a more robust indicator of effectiveness than PAND. Additionally, similar to PAND, IRD can be used to calculate Pearson's *Phi* (Φ). Parker and colleagues (2009) indicated that IRD has significant advantages including, "(a) accessible interpretation as the difference in improvement rates between baseline and treatment phases, (b) simple hand-calculations, (c) compatibility with PND from visual analysis, (d) known sampling distributions, (e) proven track record (as risk difference) in hundreds of evidence-based medical research studies, (f) few data distribution assumptions, and (g) application to complex single-case research designs and multiple data series" (p. 138). However, Parker and colleagues (2015) still caution that IRD is insensitive to data trend and should be interpreted as such.

In general, the calculation of IRD is similar to PAND, however, there is an extra step in the calculation in which the results are converted to two improvement rates, one for phase A, and one for phase B (Parker et al., 2015). IRD is calculated by first determining the improvement rate for each phase. This is conducted by comparing the number of data points demonstrating improvement with the total number of data points for that phase, where an improved data point is one in which there is no overlap with the preceding or following phase (Parker et al., 2009). Next, the ratios are used to calculate

the difference in the improvement rate from baseline to intervention (Parker et al., 2009). For the calculation of IRD, it is helpful to label the quadrants of the graph, where the baseline phase is split vertically into two sections: the improved (W), and not improved (Y) data points. Similarly, the treatment phase is split vertically into improved (X) and not improved (Z) data points (Parker et al., 2015). Improvement rates are calculated as number of improvements divided by the total number in each phase. Parker and colleagues (2009) state "an improved data point in baseline is defined as one that ties or exceeds any data point in the treatment phase; an improved data point in the treatment phase is defined as any which exceeds all data points in the baseline phase" (p. 139)

The improvement rate for baseline ($IR_A$) is calculated as W/(W+Y), and for treatment ($IR_B$) is calculated as X/(X+Z). To determine IRD, the improvement rate for the baseline phase is subtracted from the improvement rate from the treatment phase (Parker et al., 2015). The interpretation of IRD is as follows: .50 and below is a very small or questionable effect, .50 to .70 is a moderate effect, and .70 and higher is a large effect (Parker et al., 2009).

Tables 6 through 9 detail the different effect sizes across classrooms and variables.

**Research Question 1: Treatment Integrity**

For all classrooms, visual analysis reveals that there were changes in level and trend between baseline and intervention upon the introduction of performance feedback. Additionally, there is clear evidence of the immediacy of the effect for treatment integrity across classrooms. Further, PAND, *Phi*, and IRD for treatment integrity in all classrooms

was 1. In other words, none of the data points from baseline overlapped with any of the points from the treatment phase for any of the classrooms. This suggests that the introduction of PF was associated with a large and significant effect on the integrity of the class-wide behavior intervention.

The results of the visual analysis for each classroom with respect to treatment integrity are discussed below.

**Classroom 1**. In classroom 1, during the final observation in baseline, staff achieved 20% TI; however, after one session of PF, staff increased TI to 84%. Additionally, the level increased drastically from baseline to treatment phase. During baseline observations, TI ranged from 17 to 23% with an average of 20% TI. However, after starting PF, TI increased, ranging from 74 to 92% with an average of 84%. Analysis also revealed that there was little variability in TI during both phases. Finally, the data were essentially flat, and were not trending in either direction during baseline or treatment.

**Classroom 2.** Similar to classroom 1, visual analysis reveals there is clear evidence of the immediacy of the effect of performance feedback on treatment integrity. During the final observation in baseline, the staff achieved 35% TI; however, after one session of PF, staff increased TI to 62%. Additionally, the level increased drastically from baseline to treatment phase. During baseline observations, TI ranged from 19-36% with an average of 31% TI. However, after starting PF, TI increased, ranging from 62-88% with an average of 78%. Analysis also revealed that there was little variability in TI during both phases. Finally, with respect to trend and variability, the data in baseline

were essentially flat, and not trending in either direction. However, during treatment, the first two data points represented an upward trend, while the last two data points were the same at 88%.

**Classroom 3.** Similar to classrooms 1 and 2, visual analysis reveals there is clear evidence of the immediacy of the effect for treatment integrity in classroom 3. During the final observation in baseline, the staff achieved 40% TI; however, after one session of PF, staff increased TI to 70%. Additionally, the level increased drastically from baseline to treatment phase. During baseline observations, TI ranged from 29-40% with an average of 36% TI. However, after starting PF, TI increased, ranging from 70-92% with an average of 78%. Analysis also revealed that there was little variability in TI during both phases. Finally, the data were essentially flat, and were not trending in either direction during baseline or treatment.

**Classroom 4.** Similar to the other classrooms, visual analysis reveals there is clear evidence of the immediacy of the effect for treatment integrity in classroom 4. During the final observation in baseline, the staff achieved 30% TI; however, after one session of PF, staff increased TI to 60%. Additionally, the level increased from baseline to treatment phase. During baseline observations, TI ranged from 15-30% with an average of 21% TI. However, after starting PF, TI increased, ranging from 52-60% with an average of 56%.

**Research Question 2: Academic Engaged Time**

Regarding academic engaged time, visual analysis reveals that all classrooms evidenced a change in level between baseline and treatment phases. Additionally, in two

of the four classrooms (classrooms 1 and 4), there was a complete separation (i.e., no overlap of data points) between baseline and treatment. Additionally, in these two classrooms, the immediacy of the effect was noted.

With respect to quantitative analysis, the average across classrooms for each type of analysis was as follows: PAND = .87, $\Phi$ = .72, and IRD = .69. Overall, this corresponds to a large effect size. PAND for AET varied by classroom, ranging from .67 in classroom two to 1.0 in classrooms one and four (see Table 6). Additionally, $\Phi$ for each classroom ranged from .31 in classroom two to 1.0 in classrooms one and four (see Table 7). Finally, IRD ranged from .05 in classroom 2 to 1.0 in classrooms one and four. (see Table 8).

The results of the visual analysis for each classroom with respect to academic engaged time are discussed below.

**Classroom 1.** Similar to treatment integrity, the level of academic engagement increased substantially between baseline and treatment phases. The percentage of intervals in which students were engaged in baseline ranged from 38-50%; whereas during treatment, the percentage of intervals ranged from 78-100%. The average number of intervals of engagement during baseline was 20%, and the average number of intervals of engagement during treatment was 88%. In addition to changes in level and trend, there was also an immediate effect on the dependent variable once PF was introduced. During the final observation session in baseline, the percentage of intervals of student engagement was 38%, and during the first observation in the treatment condition, the percentage of intervals during which students were engaged increased to 84%.

**Classroom 2.** The percentage of intervals in which at least 80% of students were engaged during baseline ranged between 60-90%. Although the highest point in both phases was the same at 90%, the average percentage of intervals of engagement increased from 70% in baseline to 86% in treatment, signifying an increase in overall level. Additionally, although small, there was an immediate effect on the dependent variable once PF was introduced. During the final observation session in baseline, the percentage of intervals with at least 80% of students engaged was 70%, and during the first observation in the treatment condition, the percentage of intervals during which students were engaged increased to 82%.

**Classroom 3.** Similar to classroom 2, the percentage of intervals in which at least 80% of students were engaged during baseline ranged between 60-90%. The range for the treatment phase increased from baseline to 90-100%. Additionally, although there was some overlap between the phases, the average percentage of intervals of engagement increased from 77% in baseline to 95% in treatment, signifying an increase in overall level. Additionally, although small, there was an immediate effect on the dependent variable once PF was introduced. During the final observation session in baseline, the percentage of intervals with at least 80% of students engaged was 90%, and during the first observation in the treatment condition, the percentage of intervals during which students were engaged increased to 100%.

**Classroom 4.** The percentage of intervals in which at least 80% of students were engaged during baseline ranged between 0-40%. The range for the treatment phase increased slightly from baseline to 40-60%. Additionally, the average percentage of

intervals of engagement increased from 20% in baseline to 50% in treatment, signifying an increase in overall level. Additionally, there was an immediate effect on the dependent variable once PF was introduced. During the final observation session in baseline, the percentage of intervals with at least 80% of students engaged was 0%, and during the first observation in the treatment condition, the percentage of intervals during which students were engaged increased to 50%.

**Research Question 3: Disruptive Behavior**

Upon visual analysis, only one of the classrooms (Classroom 1) evidenced a change in level between the baseline and treatment phases for disruptive behavior. In all other classrooms, there was significant overlap of data points between baseline and treatment, and no change in level or trend was demonstrated.

With respect to quantitative analysis, the improvement in student disruptive behavior on average was calculated as PAND = .78, $\Phi$ = .45, and IRD = .20. For the PAND calculation, this corresponds to a large effect. However, the calculation for *phi* indicates a medium effect, and for IRD, indicates a small or negligible effect. PAND for DB varied by classroom, ranging from .67 in classroom two to 1.0 in classrooms one and four (see Table 6). Additionally, $\Phi$ for each classroom ranged from .37 in classrooms three and four to .73 in classroom one (see Table 7). Finally, IRD ranged from 0 in classrooms three and four to .53 in classroom one (See Table 8).

The results of the visual analysis for each classroom with respect to disruptive behavior are discussed below.

**Classroom 1.** Overall, a change was observed in the level of disruptive behavior between baseline and treatment conditions. The percentage of intervals in which at least one student engaged in disruptive behavior was reduced from an average of 32% in baseline to an average of 14% during treatment. In classroom 1, similar to the change observed in the other two dependent variables, the immediacy of the effect was also apparent for disruptive behavior. During the last observation session in baseline, disruptive behavior was observed during 35% of intervals; however, during the first observation in the treatment phase, disruptive behavior was only observed in 10% of intervals.

**Classroom 2.** Overall, trend and level did not change significantly between baseline and treatment for disruptive behavior. Additionally, there were more disruptive behaviors observed during treatment than during baseline; the average percentage of intervals with disruptive behavior during baseline was 6%, whereas during treatment, the average percentage of disruptive behaviors was 11%.

**Classroom 3.** Overall, level of disruptive behavior decreased slightly between baseline and treatment for disruptive behavior. The average percentage of intervals of disruptive behavior in baseline was 18%. During treatment, the average percentage of intervals decreased to 10%. The range, or variability of the data, remained the same between baseline and treatment; the baseline range was 6-30%, and the treatment range was 0-20%. Finally, there was no immediacy of the effect observed, as the last data point in baseline was the same percentage of intervals as the first data point in the treatment phase.

**Classroom 4.** Overall, level of disruptive behavior decreased slightly between baseline and treatment for disruptive behavior. The average percentage of intervals of disruptive behavior in baseline was 18%. During treatment, the average percentage of intervals decreased to 10%. The range, or variability of the data, remained the same between baseline and treatment; the baseline range was 6-30%, and the treatment range was 0-20%. Finally, there was no immediacy of the effect observed, as the last data point in baseline was the same percentage of intervals as the first data point in the treatment phase.

**Research Question 4: Social Validity**

The maximum score possible on the IRP-15 is a 6 for each question and a score of 90 overall. For each question, the participant was asked to rate each question according to a Likert scale. The Likert scale ranged from 1-6 where 1 corresponded with strongly disagree and 6 with strongly agree. With respect to the level system, analysis of the results of the IRP-15 indicates that the average treatment acceptability was 4.1 (range of 2-5) for teachers and 3.9 (range of 1-4) for instructional assistants. This corresponded with an average score of 58 points overall for both instructional assistants and teachers. Additionally, teachers and instructional assistants also rated the consultation service using performance feedback delivered by the author. Results indicate that average treatment acceptability for teachers was 4.5 (range of 2-5), and for instructional assistants was 4.2 (range of 3-5). This corresponded with an average score of 64 points overall for teachers and instructional assistants.

## Discussion

This study was designed to examine the extent to which performance feedback was associated with increases in treatment integrity to a class-wide behavior intervention for students with ED. Subsequent effects for improvement in student behavior was also measured. Assuming the accurate measurement of treatment integrity in this study, classroom staff demonstrated significant improvement in the fidelity of implementation once performance feedback was added. Before beginning performance feedback, classroom staff's delivery of the level intervention averaged 28% integrity. In other words, treatment integrity levels after one two-day training averaged 28%. This level of implementation is consistent with previous lines of research investigating teacher implementation of interventions after a day or two of professional development (Joyce & Showers, 2002). However, after performance feedback was added, staff increased their integrity to the level system to an average of 78%. Overall, results indicate that the delivery of performance feedback, consisting of the delivery of information regarding level of intervention implementation presented graphically, was associated with increases in both treatment integrity and academic engagement across four classrooms. Further, results suggest that the delivery of performance feedback impacted the level of treatment integrity after just one session across all classrooms. Moreover, these improvements were sustained over time with the continued delivery of performance feedback provided one time per week in a group consisting of teachers and instructional assistants.

Out of all of the classrooms, only the staff in one classroom did not reach a level of 80% integrity during the performance feedback phase. It is important to note, however,

that this classroom initially evidenced the lowest level of TI (20% average), and increased to an average of 60% after performance feedback began. All other classrooms demonstrated significant improvements in the delivery of the class-wide behavior intervention, achieving 80% fidelity or greater within one to two sessions after the introduction of performance feedback.

The significant improvement in TI in this study is congruent with findings from a recent synthesis identifying performance feedback as an evidence-based practice in the schools for increasing treatment integrity (Fallon et al., 2015). In this systematic review of the literature, Fallon and colleagues detailed evidence from 29 studies in which PF was delivered by consultants to increase fidelity to interventions delivered by various types of educators working with different populations of students. Based on their analysis, the majority of studies demonstrated at least a moderate effect when PF was provided. Additionally, similar to the current study, they documented that the majority of studies included verbal, graphic, and problem-solving components during the delivery of PF (Fallon et al., 2015).

. It is important to note here that in a recent meta-analysis, researchers identified that PF produced a more significant relative change in teacher behavior than student behavior (Solomon et al., 2012). These authors indicated that this finding was not surprising given that PF as an intervention is directly related to the fidelity with which the intervention is delivered, and not with how effective the intervention is itself. However, although few studies have analyzed the extent to which increases in treatment integrity to a class-wide behavior management system have also affected student outcomes, one

study did find that the implementation of a level system in one high school classroom for students with ED improved student on-task behavior (Mastropieri et al., 1998). However, in this study, treatment integrity data were not collected. As a result, questions remained regarding the extent to which the independent variable (i.e., the class-wide intervention) affected change in the dependent variable. In general, if one is not measuring the fidelity of the implementation of the independent variable, it is difficult to make claims regarding temporal precedence, and ultimately internal validity (Kratochwill et al., 2010). For this reason, in recent years, researchers have continuously emphasized the collection of fidelity data (Gresham, 2009).

Although more studies are now reporting fidelity data, many continue to question the extent to which fidelity of implementation affects participant outcomes (Durlak & DuPre, 2008; Sanetti & Kratochwill, 2009; Gresham, 2009). In other words, some have questioned if there is a threshold, whereby a certain level of fidelity in the delivery of the intervention would produce similar outcomes to those found in the literature. In this way, the findings of this study contribute to and extend the literature by demonstrating that significant increases in treatment fidelity were related to significant increases in student academic engagement. In other words, not only did this study replicate the effects of the study by Mastropieri and colleagues (1998) with respect to the implementation of a level system and subsequent improvement in student outcomes, but it also expanded upon their findings by suggesting that the implementation of performance feedback was associated with increases in student engagement. In this way, the findings were similar to Jeffrey

and colleagues (2009) in which they found a positive correlation between increases in treatment fidelity and students' time on task.

However, these same increases in treatment fidelity were not functionally related to a change in student disruptive behavior in three out of the four classrooms. In other words, better adherence to the components of the level system did not have an impact on the level of student disruptive behavior in most of the classrooms. Across all classrooms, there was overlap between the baseline and intervention phases for disruptive behavior data, revealing no improvement in this variable across conditions. Further, quantitative analysis revealed improvements in student disruptive behavior were negligible in all but one classroom. Although it is enticing to draw conclusions about improvement of disruptive behavior in the one classroom, in order to demonstrate a functional relationship between an intervention and a dependent variable in a multiple baseline design, the outcome must be replicated across conditions (Cooper et al., 2007). Although previous studies have demonstrated that increases in treatment integrity have corresponded to a decrease in student problem behavior (e.g., DiGennaro et al., 2007), the lack of improvement in student disruptive behavior in this study may be due to a number of issues. For example, in this study, during the periods in which the students were observed, some of the classrooms initially demonstrated low levels of disruptive behavior. In other words, there was a floor effect for this variable. As a result, the increases in treatment integrity in these classrooms had little to no impact on disruptive behavior. More concerning, however, is the fact that some classrooms demonstrated more significant problem behavior *after* staff began evidencing better adherence to treatment

components. Anecdotally, it appears that this may have been the result of the absence or presence of certain students with individualized behavior plans on any given day. In other words, there were certain students for whom the class-wide intervention did not appear to impact their problem behavior, and who engaged in more problem behaviors than others. As a result, these students either had individualized behavior programs or were in the process of moving to more restrictive placements. When these students were present, it appeared to artificially inflate the level of disruptive behavior for the class. Finally, this may have also been an artifact of the measurement system for disruptive behavior, since partial-interval recording was used to measure disruptive behavior, and the use of partial-interval recording can overestimate the frequency of behavior (Cooper et al., 2007). In general, however, the delivery of PF in this study was associated with improvements in both staff's treatment integrity and improvements in student academic engagement. Additionally, similar to previous studies, all staff reported that PF was a helpful, appropriate, and efficient way to provide consultation. In light of Noell and colleagues' (2005) findings comparing three consultation strategies wherein PF produced the most significant improvements in TI, PF appears to be an effective and efficient way to provide consultation in the schools. In this study, performance feedback was rated as highly as the other two consultation strategies (Noell et al., 2005).. Extrapolating from the Noell and colleagues study, in combination with the findings from this study, it appears that PF may be a socially valid way to approach consultation for staff who are implementing class-wide behavioral interventions for students with ED.

## Limitations

Several cautionary notes are worth considering concerning the limitations of this study. There are various complications related to the design used and the measures developed for this study. In general, however, it is important to note that staff in the various classrooms began receiving PF when they demonstrated a stable rate of responding in the baseline phase. These guidelines were delineated by Cooper and colleagues (2007). It is important to note that recently, some have suggested that when using a multiple baseline design, researchers should randomize the order in which intervention begins in different settings (Kratochwill & Levin, 2010). In other words, in this study, to increase internal validity, Kratochwill and Levin would have recommended that the researcher randomly selected the order in which classrooms received intervention.

**Design.** Although there are various advantages to using a multiple baseline design in single case research, there are some limitations to the types of conclusions one can draw when using this type of design. Some researchers have hypothesized that although it is the most widely used single case design, conclusions drawn from the replication of affects across settings or staff may not be as robust as those drawn from the use of a reversal design (Cooper et al., 2007). With a reversal design, the experimenter is able to manipulate the introduction and removal of the independent variable in an effort to demonstrate a functional relationship. In this way, the multiple baseline design is sometimes seen as a technique by which the dependent variable may be altered rather than a true experimental analysis (Cooper et al., 2007).

**Measurement Issues**. Due to periodic teacher absence and state testing days, scheduled observations had to be postponed on certain weeks. As a result, there were classes where on certain weeks there were no observations, and therefore no performance feedback was conducted. Research suggests that the more consistently PF is conducted, and the less amount of time between behavioral observations and delivery of PF, the more effective it is as a behavior change intervention (Solomon et al., 2012).

Additionally, two of the measures used in this study, the TIOS and the CBOT, were created by the researcher. Although pilot data was collected on the TIOS, the psychometric properties of the tool itself are unknown. As a result, it is difficult to draw conclusions regarding the validity of the tool, and the extent to which it measured the construct it purported to measure. Additionally, and perhaps more importantly, the researcher did not follow a systematic process in designing the TIOS to account for the different types of integrity discussed in the current literature. For example, according to a recent analysis by Sanetti and Collier-Meek (2014), integrity is a multidimensional construct that includes content, quantity, and process-related dimensions. The TIOS included both observational elements and a permanent product review; however, there was no consideration for the level of adherence to the plan (e.g., implemented as planned, implemented with little deviation, etc.), or the quality of the delivery (e.g., the tone of voice used to praise the students for following the rules). As a result, one could question if treatment integrity to the class-wide intervention was captured to the fullest extent possible.

With respect to the CBOT, the researcher relied heavily on the use of time-sampling or discontinuous recording procedures for measurement of student engagement and student disruptive behavior. By their nature, discontinuous recording procedures are designed to amass aggregate data that can be used to extrapolate behavior over longer periods of time. In other words, data were gathered over 20 minute time periods one day per week and were assumed to be indicative of student and staff behavior over the 30 plus hours students and staff were at school during the week. Additionally, the CBOT included two types of time-sampling methods: partial interval recording and momentary time sampling, both of which have strengths and limitations. In particular, partial interval recording, which was used to gather data on student disruptive behaviors, tends to overestimate the true occurrence of behavior (Fiske & Delmolino, 2012). Further, although momentary time sampling, used for student engagement, may represent a more accurate portrayal of student behavior, it depends upon the level and duration of the behavior being measured (Fiske & Delmolino, 2012).

In addition, the CBOT included the use of the planned activity check, a class-wide variation of momentary time sampling (Cooper et al., 2007). There are few examples of this type of observational system being used in the literature, and there is limited discussion regarding its utility. However, the alternative observation system used in some studies examining academic engagement is a discontinuous partial interval or momentary time sampling procedure, whereby one student is observed for a brief time period (e.g., 15s), data are recorded, the next student is observed, data are recorded, and so on. The theory is that by averaging the behaviors across students in the class, a conclusion can be

drawn about classroom behavior overall. This data recording system is also subject to limitations related to both averaging the data and the measurement of behavior by discontinuous recording procedures as previously discussed. Clearly, this is an area that requires future inquiry.

**Standards**. Overall, the current study only allows for consideration of meeting standards with reservations through What Works Clearinghouse because of the low number of data points in baseline (Kratochwill et al., 2010). For a study to meet WWC standards without reservations, there must be at least five data points in each phase (Kratochwill et al., 2010). Additionally, Parker and colleagues (2007) suggested that in order to conduct meaningful quantitative analysis, researchers should have at least 20 data points in each iteration of the multiple baseline design. In the current study, there were only 10 data points for TI for each classroom, thereby limiting the validity of the quantitative analysis for each classroom.

**Further concerns**. It is also important to note that improvement in TI alone may not impact student outcomes as quickly or efficiently as desired. Although this may be for any of the reasons previously discussed, it also may be a manifestation of the type of behavior intervention employed in this study. Generally, one of the limitations of this study is that the intervention used did not necessarily include the explicit teaching of new pro-social skills. Although appropriate social skills were differentially reinforced through the level system, future research in this area may want to consider the addition of a social skills program to supplement differential reinforcement. Additionally, future studies could incorporate other types of dependent variables besides disruptive behavior to help

determine the extent to which the level system improved student outcomes. Examples include the percentage of students earning the top level each week, the number of office referrals, and the number of suspensions. Further, with respect to the intervention, the researcher was familiar with the specific components of the level system, which may have influenced the way in which data were collected during observation. On a related note, during the baseline phase, the researcher was heavily involved in supporting the teachers and the general aspects of the intervention. As a result, it would be difficult to generalize the results of this study to other types of interventions that did not incorporate this level of support.

Finally, it is important to note that there are many other types of classroom management procedures that have been shown to influence student behavior. For this study, a level system was the classroom management system enlisted; however, there are many other types of interventions that can be used. Since there is no one single type of intervention that is empirically-validated to improve the behavior of students with ED, it is possible that a combination of different types of interventions would result in higher levels of academic engagement. Based on the results of this study, it is unclear if these results would generalize to other class-wide behavior management strategies. Future research could utilize performance feedback to increase treatment integrity to another type of class-wide behavioral intervention with ED. Further, the results of these studies could be compared to determine if one type of class-wide intervention was more effective than another given similar levels of treatment implementation.

Chapter 4: References

Alvero, A. M., Bucklin, B. R., & Austin, J. (2001). An objective review of the effective and essential characteristics of performance feedback in organizational settings. *Journal of Organizational Behavior Management, 21(1),* 3-31.

Baker, P. H. (2005). Managing student behavior: How ready are teachers to meet the challenge? *American Secondary Education, 33(3),* 51-64.

Barnett, D. W., Lentz, F. E., Bauer, A. M., Macmann, G., Stollar, S., & Ehrhardt, K. E. (1997). Ecological foundations of early intervention: Planned activities and strategic sampling. *The Journal of Special Education, 31(4),* 471-490.

Begeny, J.C. & Martens, B.K. (2006). Assessing pre-service teachers' training in empirically-validated behavioral instruction practices. *School Psychology Quarterly, 21(3*), 262-285.

Billingsley, B. (2004). Promoting teacher quality and retention in special education, *Journal of Learning Disabilities, 37, (5)* 370-376.

Bradley, R., Doolittle, J., & Bartolotta, R. (2008). Building on the data and adding to the discussion: The experiences and outcomes of students with disabilities. *Journal of Behavioral Education, 17(1),* 4-23.

Breton, W. (2010). Special education paraprofessionals: Perceptions of preservice preparation, supervision, and ongoing developmental training. *International Journal of Special Education, 25(1),* 24–45.

Brock, M. E. & Carter, E. W. (2015). Effects of a professional development package to prepare special education paraprofessionals to implement evidence-based practice. *The Journal of Special Education, 49(1),* 39-51.

Brossart, D.F., Parker, R.I., Olson, E.A., & Mahadevan, L. (2006). The relationship between visual analysis and five statistical analyses in a simple AB single-case research design. *Behavior Modification, 30*(5), 531-563.

Brunsting, N. C., Sreckovic, M. A., & Lane, K. L. (2014). Special education teacher burnout: A synthesis of research from 1979 to 2013. *Education and Treatment of Children, 37(4),* pp. 681-711

Bullis, M. & Cheney, D.(1999). Vocational and transition interventions for adolescents and young adults with emotional or behavior disorders. *Focus on Exceptional Children, 31(7),* 1-24.

Cancio, E. & Johnson, J. W. (2007). Level systems revisited: An important tool for educating students with emotional and behavioral disorders. *International Journal of Behavioral Consultation and Therapy, 3(4),* 512-527.

Carr E.G., Taylor J.C., & Robinson S (1991). The effects of severe problem behavior in children on the teaching behavior of adults. *Journal of Applied Behavior Analysis, 24,* 523–535.

Cataldo, M. F. & Risley, T. R. (1973). Development of a standardized measure of classroom participation. Paper Presentation from the American Psychological Association.

Cheney, D. A., Cumming, T. M., Slemrod, T. (2013). Secondary education and promising practices for students with emotional/behavioral disorders. In F. M. Gresham & H. M. Walker (Eds.) *Handbook of Evidence-Based Practices for Emotional and Behavioral Disorders* (pp. 344-360). New York: Guilford Publications.

Cohen, M.A., Piquero, A. R., & Jennings, W.G. (2010). Studying the costs of crime across offender trajectories. *Criminology and Public Policy, 9(2),* 279-305.

Cohen, M. A. & Piquero, A. R. (2009). New evidence on the monetary value of saving a high risk youth. *Journal of Quantitative Criminology*, *25*, 25-49. Doi: 10.1007/s10940-008-9057-3.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cook, B. G., Landrum, T. J., Tankersley, M. & Kauffman, J. M. (2003). Bringing research to bear on practice: Effecting evidence-based instruction for students with emotional or behavioral disorders. Education and Treatment of Children, 26(4), 345-361.

Cooper J.O, Heron T.E, & Heward, W.L. (2007). *Applied behavior analysis* (2nd ed.) Upper Saddle River, NJ: Pearson.

Cullinan, D. (2004). Classification and definition of emotional and behavioral disorders. In In R. Rutherford, M. Quinn, & S. Mathur (Eds.), *Handbook of Research in Emotional and Behavior Disorders,* (pp.32-53). New York: Guilford.

Darling-Hammond, L. & Sykes, G.. (2003). Wanted: A national teacher supply policy for education: The right way to meet the "Highly Qualified Teacher" challenge. *Education Policy Analysis Archives, 11(33).* Retrieved from: http://epaa.asu.edu/epaa/v11n33

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 (2),* 220-231

Downer, J. T., Rimm-Kaufman, S. E., & Pianta, R. C. (2007). How do classroom conditions and children's risk for school problems contribute to children's behavioral engagement in learning? *School Psychology Review, 36(3),* 413-432.

Fallon, L. M., Collier-Meek, M. A., Maggin, D. M., Sanetti, L. M., & Johnson, A. H. (2015). Is performance feedback for educations an evidence-based practice? A systematic review and evaluation based on single-case research. *Exceptional Children, 81 (2),* 227-246.

Filcheck, H. A., McNeil, C. B., Greco, L. A., & Bernard, R. S. (2004). Using a whole-class token economy and coaching of teacher skills in a preschool classroom to manage disruptive behavior. *Psychology in the Schools, 43(3*), 351-362.

Finn, C. A. & Sladeczek, I. E. (2001). Assessing the social validity of behavioral interventions: A review of treatment acceptability measures. *School Psychology Quarterly, 16(2),* 176-206.

Fiske, K. & Delmolino, L. (2012). Use of discontinuous methods of data collection in behavioral intervention: Guidelines for practitioners. *Behavior Analysis in Practice, 5(2),* 77-81.

Furlong M.J., Morrison G.M., Jimerson S.R. (2004). Externalizing behaviors of aggression and violence and the school context. In R. Rutherford, M. Quinn, & S. Mathur (Eds.), Handbook of research in behavior disorders, *Handbook of Research in Emotional and Behavioral Disorders,* (pp. 243-261). New York: The Guilford Press.

Gettinger, M., & Seibert, J.K. (2002). Best practices in increasing academic learning time. In A. Thomas (Ed.), Best practices in school psychology IV: Volume I (4th ed., pp. 773-787). Bethesda, MD: National Association of School Psychologists.

Gersten, R., Keating, T., Yovanoff, P., & Harniss, M. K. (2001). Working in special education: Factors that enhance special educators' intent to stay. *Exceptional children, 67(4),* 549-567.

Giangreco, M. F., Edelman, S. W., Broer, S. M., & Doyle, M. B. (2001). Paraprofessional support of students with disabilities: Literature from the past decade. *Exceptional Children, 68(1),* 45-63.

Giangreco, M. F., Suter, J. C., Hurley, S. M. (2014). Revisiting personnel utilization in inclusion-oriented schools. *The Journal of Special Education, 47(2),* 121-132.

Greenwood, C. R., Horton, B. T., & Utley, C. A. (2002). Academic engagement: current perspectives on research and practice. *School Psychology Review, 31(3),* 328-349.

Gresham, F. M. (2009). Evolution of the treatment integrity concept: Concurrent status and future directions. *School Psychology Review, 38(4),* 533-540.

Gresham, F. M. (2013). Measuring and analyzing treatment integrity data in research. In L. M. Sanetti & T. R. Kratochwill (Eds.) *Treatment Integrity: A Foundation for Evidence-Based Practice in Applied Psychology* (pp. 109-130). Washington DC: American Psychological Association.

Gresham, F. M. & Kern, L. (2004). Internalizing behavior problems in children and adolescents. In R. Rutherford, M. Quinn, & S. Mathur (Eds.), *Handbook of Research in Emotional and Behavior Disorders,* (pp.262-281). New York: The Guilford Press.

Gresham, F. M., Lane, K. L., MacMillan, D. L., & Bocian, K. M. (1999). Social and academic profiles of externalizing and internalizing groups: Risk factors for emotional and behavioral disorders. *Behavioral Disorders, 24(3),* 231-245.

Grieve, A. M. (2009). Teachers' beliefs about inappropriate behavior: Challenging attitudes? *Journal of Research in Special Education Needs, 9(3),* 173-179.

Hall L.J., Grundon G.S., Pope C., & Romero A.B. (2010). Training paraprofessionals to use behavioral strategies when educating learners with autism spectrum disorders across environments. *Behavioral Interventions, 25(1),* 37–51.

Henderson, K., Klein, S., Gonalez, P. & Bradley, R. (2005). Teachers of children with emotional disturbance: A national look at preparation, teaching conditions, and practices. *Behavioral Disorders, 31(1),* 6-17.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The Use of Single-Subject Research to Identify Evidence-Based Practice in Special Education. *Exceptional Children, 71,* 165–179.

Howse, R. B., Lange, G., Farran, D. C., Boyles, C. D. (2003). Motivation and self-regulation as predictors of achievement in economically disadvantaged young children. *The Journal of Experimental Education, 71(2),* 151-174.

Individuals with Disabilities Education Act, 20 U.S.C. § 1400 (2004). Retrieved from http://idea.ed.gov/explore/view/p/,root,regs,300,A,300%252E8

Jeffrey, J. L., McCurdy, B. L., Ewing, S., & Polis, D. (2009). Classwide PBIS for students with ED: Initial evaluation of an integrity tool. *Education and Treatment of Children, 32(4),* 537-550.

Jimerson, S. R., Campos, E. & Greif, J. L. (2003). Toward an understanding of definitions and measures of school engagement and related terms. *The California School Psychologist*, 8, 7-27.

Jones, V., Dohrn, E., & Dunn, C. (2004). *Creating Effective Programs for Students with Emotional and Behavioral Disorders.* Pearson: Boston.

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

Kauffman, J. M. (2001). *Characteristics of Emotional and Behavioral Disorders of Children and Youth (7th ed.).* Upper Saddle River, NJ: Prentice-Hall.

Kazdin, A. E. (1977). Assessing the clinical or applied importance of behavioral change through social validation. *Behavior Modification, 1(4), 427-451.*

Kratochwill, T. R., Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods, 15,* 124-144.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case design technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S., Rindskopf, D. M., & Shadish, W. R. (2012). Single-Case Intervention Research Design Standards. *Remedial and Special Education*, *34*(1), 26–38. doi:10.1177/0741932512452794

Landrum, T. J., Tankersley, M., & Kauffman, J. M. (2003). What is special about special education for students with emotional or behavioral disorders? *The Journal of Special Education, 37(3),* 148-156.

Lane, K. L., Wehby, J. H., Little, A., & Cooley, C. (2005). Academic, social, and behavioral profiles of students with emotional and behavioral disorders educated in self-contained classrooms and self-contained schools: Part 1 – Are they more alike than different? *Behavioral Disorders, 30(4),* 349-361.

Lane, K. L., Jolivette, K., Conroy, M. Nelson, C. M., Benner, G. J. (2011). Future research directions for the field of E/BD: Standing on the shoulders of giants. *Education and Treatment of Children, 34(4),* 423-443.

Long, A. & Maynard, B. R. (2013). Treatment integrity as adult behavior change: A review of theoretical models. In L. M. H. Sanetti & T. R. Kratochwill (Eds.), Treatment Integrity: A Foundation for Evidence-Based Practice in Applied Psychology, APA Division 16 Book Series. Washington, D.C.: APA.

Maggin, D. M., O'Keeffe, B. V., & Johnson, A. H. (2011). A Quantitative Synthesis of Methodology in the Meta-Analysis of Single-Subject Research for Students with Disabilities: 1985–2009. *Exceptionality*, *19*(2), 109–135.

Maggin, D. M., Wehby, J. H., Moore Partin, T. C., Robertson, R. & Oliver, R. M. (2011). A compariosn of the instructional context for students with behavioral issues enrolled in self-contained and general education classrooms. *Behavioral Disorders, 36(2),* 84-99.

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

Mastropieri, M.A., Jenne, T., & Scruggs, T.E. (1988). A level system for managing problem behaviors in a high school resource program. *Behavioral Disorders, 13*, 202-208.

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 (4),* 659-672.

McKissick, C., Hawkins, R. O., Lentz, F. E., Hailley, J. & McGuire, S. (2010). Randomizing multiple contingency components to decrease disruptive behaviors and increase student engagement in an urban second-grade classroom. *Psychology in the Schools, 47(9),* 944-959.

McLeskey J. & Billingsley B. (2008). How does the quality and stability of the teaching force influence the research-to-practice gap? A perspective on the teacher shortage in special education. *Remedial and Special Education, 29*, 293–305.

Merrell, K. W. & Walker, H. M. (2004). Deconstructing a definition: Social maladjustment versus emotional disturbance and moving the EBD field forward. *Psychology in the Schools, 41(8),* 899-910.

Musser, E. H., Bray, M. A., Kehle, T. J., & Jenson, W. R. (2001). Reducing disruptive behavior in students with serious emotional disturbance. *School Psychology Review, 30(2),* 294-304.

Noell, G. H. & Gansle, K. A. (2013). The use of performance feedback to improve intervention implementation in schools. In L. M. Sanetti & T. R. Kratochwill (Eds.) *Treatment Integrity: A Foundation for Evidence-Based Practice in Applied Psychology* (pp. 161-183). Washington DS: American Psychological Association.

Noell, G. H., Witt, J. C., LaFleur, L. H., Mortenson, B. P., Ranier, D. D., & LeVelle, J. (2000). Increasing intervention implementation in general education following consultation: A comparison of two follow-up strategies. *Journal of Applied Behavior Analysis, 33,* 271-284.

Oliver, R. M. & Reschly, D. J. (2010). Special education teacher preparation in classroom management: Implications for students with emotional and behavioral disorders. *Behavioral Disorders, 35(3),* 188-199.

Parker, R. I., Hagan-Burke, S. (2007). Useful effect size interpretations for single case research. *Behavior Therapy, 38,* 95 – 105.

Parker, R. I., Hagan-Burke, S., Vannest, K. (2007). Percent of all non-overlapping data (PAND): An alternative to PND. *The Journal of Special Education, 40,* 194 – 204.

Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single-case design. *Exceptional Children, 75(2),* 135-150.

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine non-overlap techniques. *Behavior Modification, 35(4),* 303-332.

Parker, R. I., Vannest, K. J., & Davis, J. L. (2015). Non-overlap analysis for single case design. In T. R. Kratochwill & J. R. Levin (Eds.) *Single-Case Intervention Research* (pp. 127-152). Washington, DC: American Psychological Association.

Piersal, W. C. & Gutkin, T. B. (1983). Resistance to school-based consultation: A behavioral analysis of the problem. *Psychology in the Schools, 20,* 311-320.

Quinn, M. M. & Poirier, J. M. (2004). Linking prevention research with policy: Examining the costs and outcomes of the failure to prevent emotional and behavioral disorders. In R. B. Rutherford, M. M. Quinn, & S. R. Mathur (Eds.) *Handbook of Research in Emotional and Behavioral Disorders.* The Guildford Press: New York.

Quinn, M. M., Rutherford, R. B., Leone, P. E., Osher, D. M., & J. M. (2005). Youth with disabilities in juvenile corrections: A national Survey. *Exceptional Children, 71(3),* 339-345.

Sanetti, L. M. H. & Collier-Meek, M. A. (2014). Increasing the rigor of procedural fidelity assessment: An empirical comparison of direct observation and permanent product review methods. *Journal of Behavioral Interventions, 23*, 60-88.

Sanetti, L. M. H. & Kratochwill, T. R. (2009). Treatment integrity assessment in the schools: An evaluation of the treatment integrity planning protocol. *School Psychology Quarterly, 24(1),* 24-15.

Sanetti, L. M. H., Kratochwill, T. R., & Long, A. C. J. (2013). Applying adult behavior change theory to support mediator-based intervention implementation. *School Psychology Quarterly, 28(1),* 47-62.

Scheeler, M. C. (2008). Generalizing effective teaching skills: The missing link in teacher preparation. *Journal of Behavioral Education, 17 (2),* 145-159.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Statistical conclusion validity and internal validity. *Experimental and quasi-experimental designs for generalized causal inference,* 45-48.

Shapiro, E. S. (2004). Direct observation: manual for the Behavioral Observation of Students in Schools (BOSS).

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. Education and Treatment of Children, 31, 351– 380

Solomon, B. G., Klein, S. A., & Politylo, B. C. (2012). The effect of performance feedback on teachers' treatment integrity: A meta-analysis of the single-case literature. *School Psychology Review, 41 (2),* 160.

Stockall, N. S. (2014). When an aide really becomes an aid: Providing professional development for special education paraprofessionals. Teaching Exceptional Children, 46(6), 197-205.

Sutherland, K. S., Lewis-Palmer, T., Sticheer, J., & Morgan, P. L. (2008). Examining the influence of teacher behavior and classroom context on the behavioral and academic outcomes for students with emotional or behavioral disorders. *The Journal of Special Education, 41(4), 223-233.*

Sutherland, K. S. & Oswald, D. P. (2005). The relationship between teacher and student behavior in classrooms for students with emotional and behavioral disorders: Transactional processes. *Journal of Child and Family Studies, 14(1),* 1-14.

Sutherland, K. S. & Wehby, J. H. (2001). The effect of self-evaluation on teaching behavior in classrooms for students with emotional and behavioral disorders. *The Journal of Special Education, 35(3),* 161-171.

Sutherland, K. S., Wehby, J. H., & Copeland, S. R. (2000). Effect of varying rates of behavior specific praise on the on-task behavior of students with EBD. *Journal of Emotional and Behavioral Disorders, 8,* 2–8.

Tankersley, M., Landrum, T., & Cook, B. G. (2004). How research informs practice in the field of emotional and behavioral disorders. In R. B. Rutherford, M. M. Quinn, & S. R. Mathur (Eds.) *Handbook of Research in Emotional and Behavioral Disorders.* The Guildford Press: New York.

Telzrow, C. F., & Beebe, J. J. (2002). Best practices in facilitating intervention adherence and integrity. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., pp. 503–516). Bethesda, MD: National Association of School Psychologists.

U.S. Department of Education, National Center for Education Statistics (2012). *Digest of Education Statistics, 2011* (NCES 2012-001). Retrieved from: http://nces.ed.gov/fastfacts/display.asp?id=64

U.S. Department of Education, Office of Special Education Programs (2013). *Table 204.60: Percentage distribution of students 6 to 21 years old served under IDEA. R*etrieved from http://tadnet.public.tadnet.org/pages/712.

Wagner, M., Kutash, K., Duchnowski, A. J., Epstein, M. H., & Sumi, W. C. (2005). The children and youth we serve: A national picture of the characteristics of students with emotional disturbances receiving special education. *Journal of Emotional and Behavioral Disorders, 13(2),* 79-96.

Wehby, J. H., Symons, F. J., & Canale, J. A. (1998). Teaching practices in classrooms for students with emotional and behavioral disorders: Discrepancies between recommendations and observations. *Behavioral Disorders, 24(1),* 51-56.

Westover, J. M. & Martin, E. J. (2014). Performance feedback, paraeducators, and literacy instruction for students with significant disabilities. *Journal of Intellectual Disabilities, 18(4),* 364-381.

Wickstrom, K. F., Jones, K. M., LaFleur, L. H. & Witt, J. C. (1998). An analysis of treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 13*, 141-154.

Witt, J. C. (1986). Teachers' resistance to the use of school-based interventions. *Journal of School Psychology, 24*, 37-44.

Witt, J. C., VanDerHeyden, A. M., & Gilbertson, D. (2004). Troubleshooting behavioral interventions. A systematic process for finding and eliminating problems. School Psychology Review, 33, 363-383.

Table 1

*Components of Level System Targeted for Measurement of Treatment Integrity*

| Domain | Specific Task | Measurement Method |
|---|---|---|
| Staff Use of Behavioral Strategies | Ignoring low levels of inappropriate behavior/appropriate use of behavioral correction procedures | Direct Observation |
| | Behavior specific praise awarded | Direct Observation |
| | Awarding points with behavior specific praise | Direct Observation |
| | At least 5 opportunities for students to earn points each hour | Direct Observation/ Permanent Product |
| | Staff mark points on data sheet after providing behavior specific praise | Direct Observation |
| | 4:1 ratio of praise to corrective feedback | Direct Observation |
| Materials | Four to eight class rules posted and clear definitions of exactly that which is expected | Permanent Product |
| | Posted levels with descriptors of what students have access to with each level | Permanent Product |
| | Student names posted under levels | Permanent Product |
| | Level privileges evident in the class environment (i.e., board games, mechanical pencils, etc. | Direct Observation |
| Teaching Behavioral Expectations | Use of Behavioral Scripts | Direct Observation |
| | Appropriate use of level system | Direct Observation/ Permanent Product |

Table 2

*Sample Level Privileges in One Secondary Classroom*

| Level | Privileges Awarded | |
|---|---|---|
| Level 1 | Restricted seating<br>Supervised lunch | Supervised bathroom breaks |
| Level 2 | Unsupervised lunch<br>Change date on white board<br>Feed fish | Used pencil to borrow for the period<br>Conditional bathroom time<br>Assigned seating |
| Level 3 | TA duties<br>Free drawing or reading at desk after work is finished<br>Restroom pass | Plus, choose 1: snack (1 per day, positive note or call home, early lunch, cell phone charging privileges) |
| Level 4 | Cell phone charging privileges<br>Borrow mechanical pencil<br>Early lunch<br>Preferential seating | Use of technology after task completion and at lunch<br>Plus, choose 1: request off campus lunch, homework pass, outside/free time after task completion |

*Note.* Level 1 is most restrictive and level 4 is least restrictive. Levels are changed daily based on the points earned by the student the previous day.

Table 3

*Treatment Integrity Data*

| Class | Baseline Range | Baseline Average | Last Data Point Baseline | First Data Point Treatment | Treatment Range | Treatment Average |
|---|---|---|---|---|---|---|
| 1 | 17-20 | 20 | 20 | 84 | 74-92 | 84 |
| 2 | 19-36 | 31 | 35 | 62 | 62-88 | 78 |
| 3 | 29-40 | 36 | 40 | 70 | 70-92 | 78 |
| 4 | 15-30 | 21 | 30 | 60 | 52-60 | 56 |

Table 4

*Academic Engagement Data*

| Class | Baseline Range | Baseline Average | Last Data Point Baseline | First Data Point Treatment | Treatment Range | Treatment Average |
|---|---|---|---|---|---|---|
| 1 | 38-50 | 45 | 38 | 84 | 78-100 | 88 |
| 2 | 60-90 | 70 | 70 | 82 | 82-90 | 86 |
| 3 | 60-90 | 77 | 90 | 100 | 90-100 | 95 |
| 4 | 0-40 | 20 | 0 | 50 | 40-60 | 48 |

Table 5

*Disruptive Behavior Data*

| Class | Baseline Range | Baseline Average | Last Data Point Baseline | First Data Point Treatment | Treatment Range | Treatment Average |
|---|---|---|---|---|---|---|
| 1 | 26-35 | 32 | 35 | 10 | 5-30 | 14 |
| 2 | 3-10 | 6 | 5 | 19 | 3-19 | 11 |
| 3 | 6-30 | 18 | 10 | 10 | 0-20 | 10 |
| 4 | 23-100 | 59 | 100 | 50 | 20-50 | 34 |

Table 6

*PAND Across Classrooms and Variables*

| Classroom | Treatment Integrity | Academic Engaged Time | Disruptive Behavior |
|---|---|---|---|
| 1 | 1 | 1 | 1 |
| 2 | 1 | .67 | .67 |
| 3 | 1 | .80 | .70 |
| 4 | 1 | 1 | .70 |

Table 7

*Phi across Classrooms and Variables*

| Classroom | Treatment Integrity | Academic Engaged Time | Disruptive Behavior |
|---|---|---|---|
| 1 | 1 | 1 | .73 |
| 2 | 1 | .31 | .31 |
| 3 | 1 | .58 | .37 |
| 4 | 1 | 1 | .37 |

Table 8

*Improvement Rate Difference across Classrooms and Variables*

| Classroom | Treatment Integrity | Academic Engaged Time | Disruptive Behavior |
|---|---|---|---|
| 1 | 1 | 1 | .53 |
| 2 | 1 | .05 | .25 |
| 3 | 1 | .17 | 0 |
| 4 | 1 | 1 | 0 |

Table 9

*Average Effect Sizes across Variables*

| | PAND | *Phi* | IRD | Effect |
|---|---|---|---|---|
| Treatment Integrity | 1 | 1 | 1 | Large |
| Academic Engaged Time | .87 | .72 | .55 | Moderate |
| Disruptive Behavior | .77 | .44 | .20 | Small/Negligible |

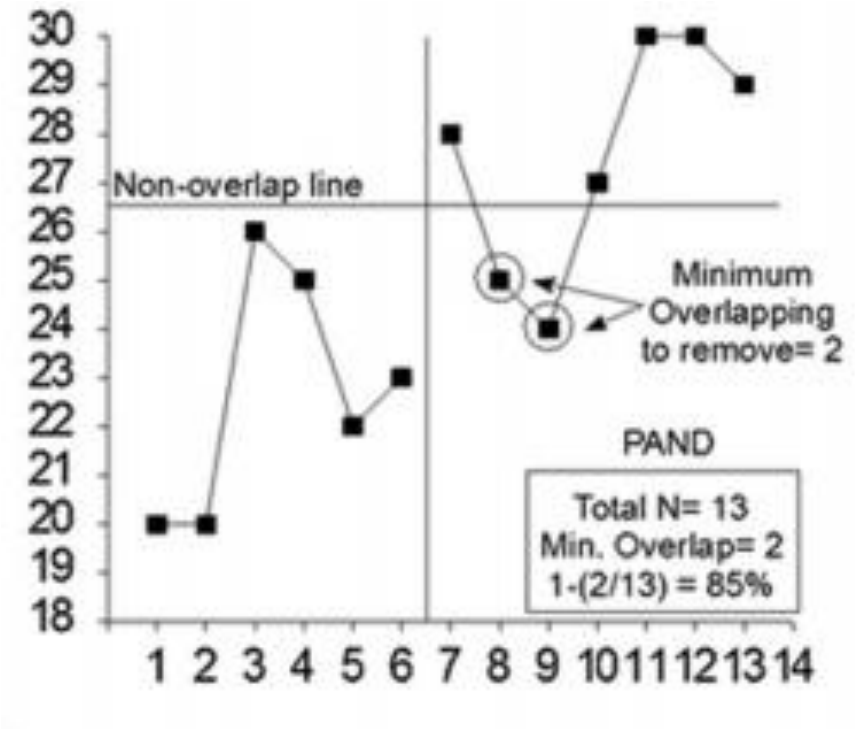


*Figure 1*. PAND Example from Parker et al., 2007

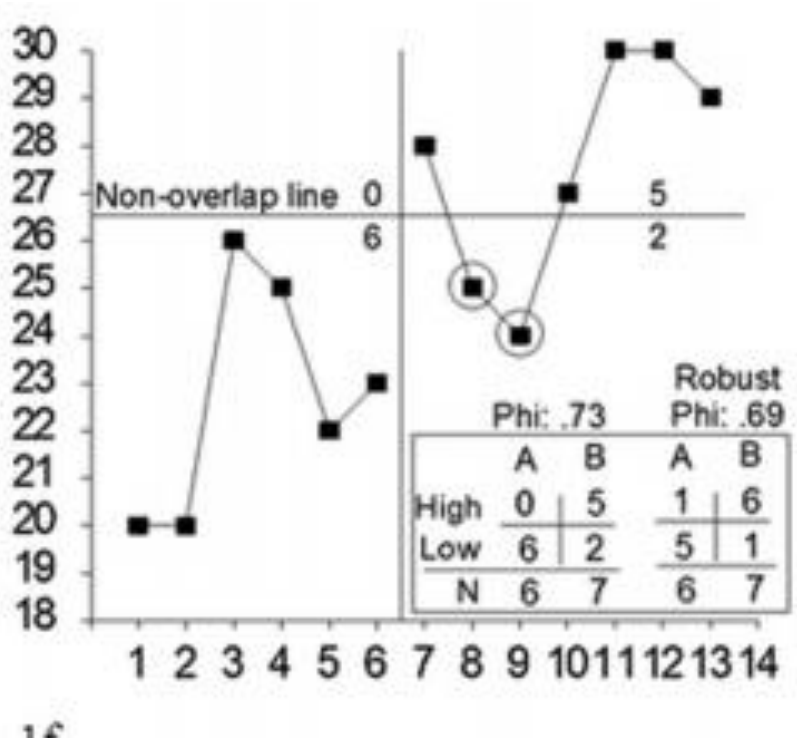


*Figure 2*. Phi Example from Parker et al., 2007

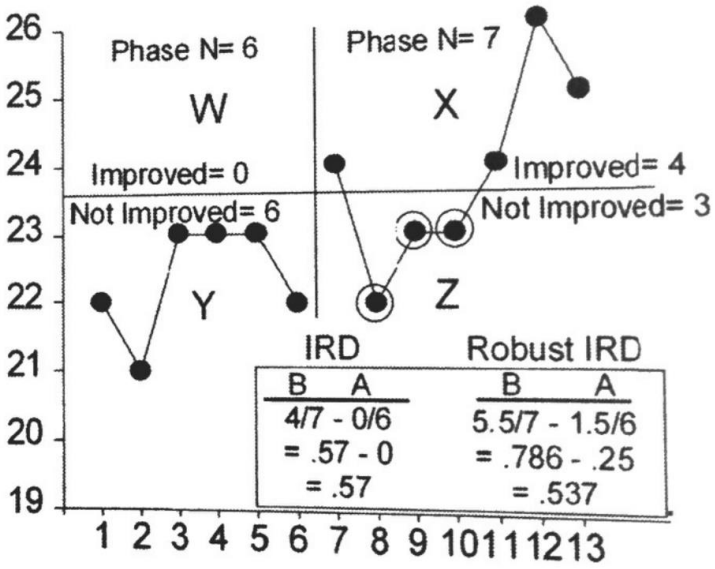


*Figure 3*. IRD Example from Parker et al., 2009

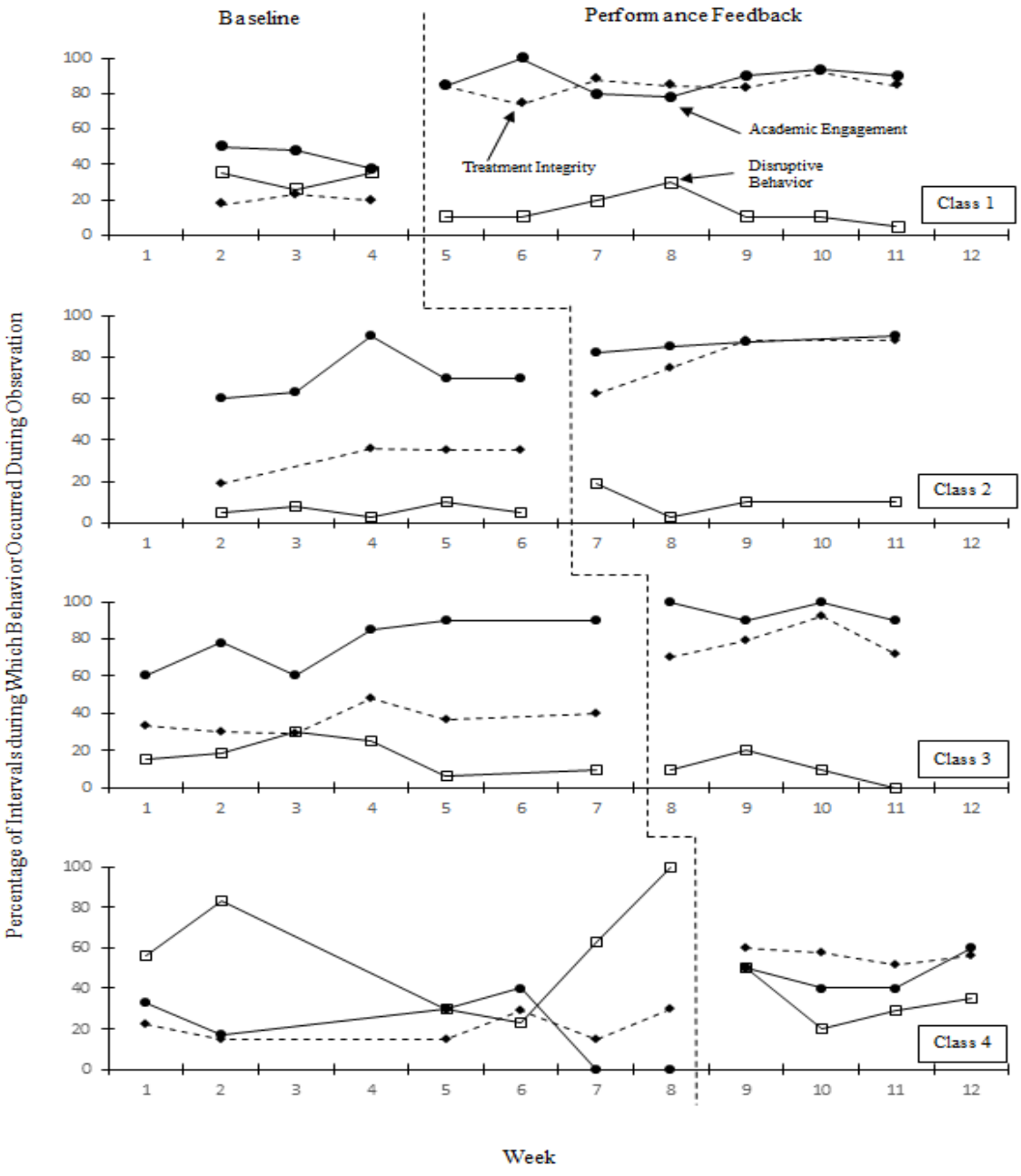


*Figure 4*. Graphed Results of Outcome Variables

*Appendix A.* Treatment Integrity Observation System (TIOS)

School:_____ Date of observation:_____

Teacher:_____ Grade:_____

Subject area:_____

Observation start time:_____ End time:_____

IOA: yes no IOA observer:_____

---

**Permanent Product Review**

**Materials (room environment)**

| | Yes | No | N/A | Comments | Score (yes=1; no=0) |
|---|---|---|---|---|---|
| Four to eight class rules posted and clear definitions of exactly that which is expected | | | | | |
| Posted levels with descriptors of what students have access to with each level | | | | | |
| Student names posted under levels | | | | | |

**Materials (collected from teacher)**

| | | | | | |
|---|---|---|---|---|---|
| Staff-developed behavioral scripts | | | | | |
| At least 5 points awarded per hour | | | | | |
| Points awarded for work completion | | | | | |
| Communication system for sending information from system home | | | | | |

**-End Permanent Product Review-**

## Direct observation (Part 1)

| | | Yes | No | N/A | Score (yes=1; no=0) |
|---|---|---|---|---|---|
| **Behavior Scripts** | Staff use behavioral scripts to teach expected behaviors | | | | |
| | Staff incorporates student role play into teaching expectations | | | | |
| **Level System** | Students use/access items listed on their corresponding level throughout class period (not just at one point in period) | | | | |
| | Level privileges are evident in the class environment (i.e., board games, mechanical pencils, etc. | | | | |
| | | **Box 2: Direct observation part 1 Total** | | | |

## Direct Observation (Part Two - 15 minute observation/30 second partial interval)

Instructions: In opportunity box at top of form, mark "X" if any student in class exhibits behavior defined in classroom expectations; in opportunity box in middle of form, mark "X" if any student in class exhibits minor off-task behavior (defined in manual). The rest of the form is completed as follows:

Top box in each cell: Correct: + Incorrect: -

Behavior Specific Praise and Awarding of Points

| Interval | :30 | 1:00 | 1:30 | 2:00 | 2:30 | 3:00 | **A** | 3:30 | 4:00 | 4:30 | 5:00 | 5:30 | 6:00 | **B** | 6:30 | 7:00 | 7:30 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Opportunity | | | | | | | 4 | | | | | | | 4 | | | |
| 1. Staff verbally provide praise to student | | | | | | | | | | | | | | | | | |
| 2. Verbally presented praise is behavior specific | | | | | | | | | | | | | | | | | |
| 3. Staff mark corresponding behavior on data sheet | | | | | | | | | | | | | | | | | |
| *Score (only enter "1" if staff engaged in all 3 behaviors)* | | | | | | | | | | | | | | | | | |

Use of Behavior Correction Procedure (used when students engage in minor off-task behaviors)

| | | | | | | | **C** | | | | | | | **D** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. staff give verbal directive to engage in appropriate bx and wait 5 seconds | | | | | | | | | | | | | | | | | |
| 2. staff again prompt student to engage in appropriate behavior and wait 5 seconds | | | | | | | | | | | | | | | | | |
| 3. Staff tell student they are on time out from points for a specified period of time | | | | | | | | | | | | | | | | | |
| 4. Staff make corresponding time out from points mark on data sheet | | | | | | | | | | | | | | | | | |
| 5. During time out from points, staff continue to provide praise | | | | | | | | | | | | | | | | | |

Shaded boxes = Opportunity

Behavior Specific Praise and Awarding of Points

| Interval | 8:00 | 8:30 | 9:00 | E | 9:30 | 10:00 | 10:30 | 11:00 | 11:30 | 12:00 | F | 12:30 | 13:00 | 13:30 | 14:00 | 14:30 | 15:00 | G |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Opportunity | | | | 4 | | | | | | | 4 | | | | | | | 4 |
| 1. Staff verbally provide praise to student | | | | | | | | | | | | | | | | | | |
| 2. Verbally presented praise is behavior specific | | | | | | | | | | | | | | | | | | |
| 3. Staff mark corresponding behavior on data sheet | | | | | | | | | | | | | | | | | | |
| *Score (only enter "1" if staff engaged in all 3 behaviors)* | | | | | | | | | | | | | | | | | | |

Use of Behavior Correction Procedure (used when students engage in minor off-task behaviors)

| | | | | H | | | | | | | I | | | | | | | J |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. staff give verbal directive to engage in appropriate bx and wait 5 seconds | | | | | | | | | | | | | | | | | | |
| 2. staff again prompt student to engage in appropriate behavior and wait 5 seconds | | | | | | | | | | | | | | | | | | |
| 3. Staff tell student they are on time out from points for a specified period of time | | | | | | | | | | | | | | | | | | |
| 4. Staff make corresponding time out from points mark on data sheet | | | | | | | | | | | | | | | | | | |
| 5. During time out from points, staff continue to provide praise | | | | | | | | | | | | | | | | | | |

**--End Observation**

Scoring for direct observation part 2 – for *behavior specific praise (top section)*, if any of the preceding opportunity boxes are marked, the number 4 goes in the box under the letter (A-J). If, in any column, numbers 1-3 are observed, place a score of 1 in the bottom box of that column. Under the letters in each column, if there are any observations for numbers 1-3, place a 1 in the corresponding box. Additionally, if there are any bottom boxes filled in with a 1, place a 1 in the bottom box under the letter. For *Use of behavior correction procedure (bottom section)*, if any of the opportunities are checked in the preceding boxes, place a 1 in the box under the letter. If, however, the behavior continues and staff continue to engage in #s 1-5, the number of opportunities should reflect this. If staff engage in #1-4, then the number 4 goes in the box and staff receive points for each of the 4 behaviors.

**Praise Score**

| Praise: | Corrective feedback: |
|---|---|
| Ratio:_____/_____ = _____<br># of praise statements # CF statements | 4:1 ratio satisfied: Yes No |
| Score Criteria:<br>4:1 ratio = 4; 2:1 ratio = 2; 1:1 ratio = 0; 1:1+ ratio = -1: | Score:<br>_____ |

**TI Score***

| | Total Correct (TC) | Total Opportunities (TO) |
|---|---|---|
| 1. Materials/Permanent Product Review (from page 1) – Box 1 | | 6 or 7 |
| 2. Direct Observation Part 1 (from page 1) – Box 2 | | 2 3 or 4 |
| 3. Use of Behavioral Strategies (add boxes A-J; pages 2 and 3) | | |
| 4. Praise Score (from above) | | 4 |
| TOTAL (add columns) | | |

**Total Treatment Integrity**

_____/_____x 100 = _____
Total Correct Total Opportunities

## Interobserver Agreement (percentage agreement)

| | Observer 1 Score | Observer 2 Score | % Agreement |
|---|---|---|---|
| Materials/Permanent Product Review (from page 1) – Box 1 | | | |
| Direct Observation Part 1 (from page 1) – Box 2 | | | |
| Use of Behavioral Strategies (add boxes A-J; pages 2 and 3) | | | |
| Praise Score (from above) | | | |
| TOTAL (add columns) | | | |

*Appendix B.* Class-wide Behavior Observation Tool (CBOT)

School:_____ Date of observation:_____

Teacher:_____ Grade:_____

Subject area:_____

Observation start time:_____ End time:_____

IOA: yes no IOA observer:_____

---

Observation Codes

| Type | Definition |
|---|---|
| Academic Engagement | Student is demonstrating any of the following behaviors: using classroom materials according the lesson, writing, reading, raising hand, talking to the teacher about assigned material, talking to a peer about the assigned material, sitting in assigned seat/area |
| Disruptive | Disruptive behavior includes any behavior that causes the teacher or other students to look in the student's direction, interrupts the teacher or other students due to audible noise or movement, or involves the student physically or verbally engaging with another person in the room that removes them from engaging in the task. Examples include making audible noises, calling out, making unauthorized comments, hitting others, throwing objects |

**Time Sampling Recording Sheet**

Academic engagement (AE) and off-task (OT) behaviors are recorded every 30 seconds via a *momentary time sampling* procedure. In this procedure observers start by looking at the left side of the room and then scan the room from left to right, recording the number of students who are **off-task** (the spot for recording during observation is underlined).

Disruptive behaviors are recorded via *partial interval recording* every 15 seconds. For each 15 second interval, each student engaging in disruptive behavior should be indicated with a tick mark. Do not count the same student twice in one interval.

Number of students in class during observation period:_____

| **Momentary Time Sampling – off-task and academic engagement** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| :30 | | 1:00 | | 1:30 | | 2:00 | | 2:30 | | 3:00 | |
| OT | AE | OT | AE | OT | AE | OT | AE | OT | AT | OT | AE |
| ____ | | ____ | | ____ | | ____ | | ____ | | ____ | |
| **Partial Interval Recording – disruptive behaviors** | | | | | | | | | | | |
| 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s |
| | | | | | | | | | | | |

| **Momentary Time Sampling – off-task and academic engagement** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3:30 | | 4:00 | | 4:30 | | 5:00 | | 5:30 | | 6:00 | |
| OT | AE | OT | AE | OT | AE | OT | AE | OT | AT | OT | AE |
| ____ | | ____ | | ____ | | ____ | | ____ | | ____ | |
| **Partial Interval Recording – disruptive behaviors** | | | | | | | | | | | |
| 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s |
| | | | | | | | | | | | |

| **Momentary Time Sampling – off-task and academic engagement** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 6:30 | | 7:00 | | 7:30 | | 8:00 | | 8:30 | | 9:00 | |
| OT | AE | OT | AE | OT | AE | OT | AE | OT | AT | OT | AE |
| ____ | | ____ | | ____ | | ____ | | ____ | | ____ | |
| **Partial Interval Recording – disruptive behaviors** | | | | | | | | | | | |
| 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s |
| | | | | | | | | | | | |

| **Momentary Time Sampling – off-task and academic engagement** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 9:30 | | 10:00 | | 10:30 | | 11:00 | | 11:30 | | 12:00 | |
| OT | AE | OT | AE | OT | AE | OT | AE | OT | AT | OT | AE |
| ____ | | ____ | | ____ | | ____ | | ____ | | ____ | |
| **Partial Interval Recording – disruptive behaviors** | | | | | | | | | | | |
| 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s |
| | | | | | | | | | | | |

| **Momentary Time Sampling – off-task and academic engagement** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 12:30 | | 13:00 | | 13:30 | | 14:00 | | 14:30 | | 15:00 | |
| OT | AE | OT | AE | OT | AE | OT | AE | OT | AT | OT | AE |
| ____ | | ____ | | ____ | | ____ | | ____ | | ____ | |
| **Partial Interval Recording – disruptive behaviors** | | | | | | | | | | | |
| 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s | 15s |
| | | | | | | | | | | | |

Scoring

| | Disruptive Behavior | All Students Academically Engaged | 1 Student Off-Task | 2 Students Off-Task | 3+ Students Off-Task |
|---|---|---|---|---|---|
| Number of Intervals | | | | | |
| Percentage of Intervals | | | | | |

IOA Scoring

| | % Intervals - Disruptive Behavior | % Intervals - All Students Academically Engaged | % Intervals - 1 Student Off-Task | % Intervals - 2 Students Off-Task | % Intervals - 3+ Students Off-Task |
|---|---|---|---|---|---|
| Observer 1 | | | | | |
| Observer 2 | | | | | |
| % Agreement | | | | | |

Appendix C. Intervention Rating Profile – 15 (IRP-15)

## Intervention Rating Profile –15 (IRP-15)

The purpose of this questionnaire is to obtain information that will aid in the selection of classroom interventions. These interventions will be used by teachers of children with behavior problems. Please circle the number that best describes your agreement or disagreement with each statement using the scale below.

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

| | | |
|---|---|---|
| 1. | This would be an acceptable intervention for the child's problem behavior. | 1 2 3 4 5 6 |
| 2. | Most teachers would find this intervention appropriate for behavior problems in addition to the one described. | 1 2 3 4 5 6 |
| 3. | This intervention should prove effective in changing the child's problem behavior. | 1 2 3 4 5 6 |
| 4. | I would suggest the use of this intervention to other teachers. | 1 2 3 4 5 6 |
| 5. | The child's problem behavior is severe enough to warrant use of this intervention. | 1 2 3 4 5 6 |
| 6. | Most teachers would find this intervention suitable for the behavior problem described. | 1 2 3 4 5 6 |
| 7. | I would be willing to use this intervention in the classroom setting. | 1 2 3 4 5 6 |
| 8. | This intervention would *not* result in negative side effects for the child. | 1 2 3 4 5 6 |
| 9. | This intervention would be appropriate for a variety of children. | 1 2 3 4 5 6 |
| 10. | This intervention is consistent with those I have used in classroom settings. | 1 2 3 4 5 6 |
| 11. | The intervention was a fair way to handle the child's problem behavior. | 1 2 3 4 5 6 |
| 12. | This intervention is reasonable for the behavior problem described. | 1 2 3 4 5 6 |
| 13. | I liked the procedures used in this intervention. | 1 2 3 4 5 6 |
| 14. | This intervention was a good way to handle this child's behavior problem. | 1 2 3 4 5 6 |
| 15. | Overall, this intervention would be beneficial for the child. | 1 2 3 4 5 6 |

*Appendix D.* Procedural Integrity Checklist

Date:_____

Classroom:_____

PF #:_____

| Step | Description of Step | Completed: |
|---|---|---|
| 1 | Greet staff members | |
| 2 | Provide most recent TI graph for all members to review (1 per staff member) | |
| 3 | Verbally provide information, while referencing graph, on percentage of integrity achieved for each component measured from the most recent observation period | |
| 4 | Provide all staff members with graphs measuring their TI progress over time | |
| 5 | Verbally provide information, while referencing graph, on average percentage of integrity over time | |
| 6 | Set goals for next meeting | |
| 7 | Verbally answer all questions from staff | |
| | Meeting duration less than 15 minutes | |
| | | Total PI:<br>____/8*100 =<br>_____ |

IRR for PI: _____/_____*100 = _____