UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Performance Feedback to Improve Classroom Management Practices

Permalink

https://escholarship.org/uc/item/4vk9m5p5

Author

Garcia, Melissa

Publication Date

2015

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Performance Feedback to Improve Classroom Management Practices

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

Doctor of Philosophy

in

Education

by

Melissa Jeanne Garcia

June 2015

Dissertation Committee:
Dr. Michael Vanderwood, Chairperson
Dr. Cathleen Geraghty
Dr. Cixin Wang

Copyright by
Melissa Jeanne Garcia
2015

The Dissertation of Melissa Jeanne Garcia is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

For my parents, husband, and Ava whose love, support, patience, and constant encouragement gave me the strength to finish strong

ABSTRACT OF THE DISSERTATION

Performance Feedback to Improve Classroom Management Practices

by

Melissa Jeanne Garcia

Doctor of Philosophy, Graduate Program in Education
University of California, Riverside, June 2015
Dr. Michael Vanderwood, Chairperson

Performance feedback has been beneficial at enhancing the treatment integrity of behavioral intervention outcomes. However, few researchers have attempted to use performance feedback to increase teachers' use of classroom management practices. The purpose of this study was to determine if there is a functional relationship between performance feedback and treatment integrity of targeted classroom management practices. In addition, the extent to which performance feedback can enhance student behavioral outcomes was examined. Direct observation data were used to report teacher use of targeted classroom management practices, student engagement, and student disruptive behavior. The functional relationship between performance feedback and teacher implementation fidelity of classroom management practices and student behavioral outcomes was examined using a single-case multiple baseline design. Participants included five teachers from an urban school district in Southern California. Once a week, for 10-weeks, teachers met for performance feedback sessions of classroom

management practices and student engagement with the consultant. Effect size analyses were used in conjunction with visual analyses. Results showed that a functional relationship between performance feedback and improved fidelity of two classroom management practices was obtained (specific praise and error correction), as well as increased engagement and decreased disruptive behavior. This study provides evidence that performance feedback can be useful in improving student engagement and reducing disruptive behavior at a classwide level. Additionally, teachers found performance feedback to be socially valid. Implications and suggestions for future research are also discussed.

Table of Contents

List of Tables

List of Figures

Performance Feedback to Improve Classroom Management Practices

Many teachers are faced with challenging disruptive student behavior that impacts their ability to teach (Markow, Moessner, & Horowitz, 2006). Managing disruptive classroom behavior reduces the amount of time teachers devote to instruction (Reinke, Lewis-Palmer, & Martin, 2007) and poor classroom management has been tied to long-term negative academic, behavioral, and social student outcomes (Kellam, Ling, Merisca, Brown, & Ialongo, 1998; Reinke & Herman, 2002). Moreover, many pre-service teacher training programs do not adequately prepare teachers in classroom management practices (Begeny & Martens, 2006), and traditional models of professional development (e.g., training without follow-up) are often ineffective (Fixsen, Naoom, Blasé, Freidman, & Wallace, 2005; Kinkead, 2007). However, one professional development technique that has been found to be effective at enhancing the treatment integrity of behavioral interventions is performance feedback. Unfortunately, few studies have examined the relationship between performance feedback with a teacher and the fidelity of implementation of evidence-based classroom management practices. In addition, those studies that have examined performance feedback and classroom management have failed to answer lingering questions within the performance feedback literature. Therefore, the present study will address this gap by exploring still unanswered questions in the performance feedback area. This will be achieved through conducting training similar to traditional behavioral consultation, following a standardized protocol for performance feedback (Fallon, Collier-Meek, Maggin, Sanetti, & Johnson, 2015; Noell et al., 2005), and providing weekly performance feedback using graphs of teacher and student data.

This research will demonstrate whether performance feedback with a teacher can decrease disruptive student behavior and increase student academic engagement as well as improve the treatment integrity of teachers' use of evidence-based classroom management practices. This paper will highlight the need for increased use of performance feedback specifically targeting evidence-based classroom management practices, which can alleviate disruptive student behavior and lead to gains in academic engagement.

Outcomes of Disruptive Student Behavior

A growing number of students are displaying disruptive behaviors within the classroom (Nelson, 1996). Such behaviors have become one of the most important issues facing schools as students and teachers report they are seriously concerned about their safety at school (National Association of State Boards of Education, 1997). To address these behaviors, an alarming number of schools are suspending students at a rate of 3.3 million students a year (Skiba et al., 2006). The majority of these suspensions are for minor misbehavior, including disruptive and insubordinate behavior. Based on a 2006 review of exclusionary and zero-tolerance disciplinary policies, the American Psychological Association (APA) found no evidence that the use of suspension, expulsion, or zero-tolerance policies resulted in improved student behavior or increases in school safety (Skiba et al., 2006). Disruptive students also experience many adverse outcomes during their education, including loss of time for academic lessons, referral to alternative education programs, expulsion, and dropping out of school (Lewis, 2001). Not only do disruptive students detract from their own learning, they also impede the learning

of classmates (Lewis, 2001), whereby 56% of students reported that distractions by other students detract from their learning (U.S. Department of Education [USDOE], 2005). Students have also reported that their teachers allot more time disciplining disruptive students than they do teaching (Johnson, Duffet, Vine, & Moye, 2003).

Disruptive student behavioral patterns. Aggressive, disruptive behavior has also been an important predictor of adolescent and adult drug use, conduct disorders, antisocial personality disorder, and criminal behavior (Dishion et al., 1996; Kellam et al., 1998; Patterson et al., 1992). In the preschool years, this early risk factor is most likely to be found in the interactions with children and their parents in the home setting. Parenting behaviors have been consistently identified as coercive, irritable, and ineffective in the development of conduct problems throughout childhood (Patterson et al., 1992). Patterson and colleagues (1992) have proposed an interactional perspective that views early-onset and chronic antisocial behavior as the outcome of coercive parenting practices. Following this theory, antisocial behavior patterns in children are likely to arise when parents use harsh, punitive, and inconsistent parenting practices instead of clear, firm, but warm responses when children display inappropriate behavior. A pattern of coercive behaviors from both parents and child develops, whereby parents withdraw requests and adhere to the child's escalating demands. In response, parents use harsh and abusive discipline practices when the child escalates. The parent's punishment is reinforced when the child temporarily capitulates, providing mutual training for the child's inappropriate behavior and harsh discipline on the parent's part. A chain of events develops: (1) the parent ignores or attacks the child with angry or aggressive actions, (2)

the child counterattacks with anger or aggression, (3) in response the parent disengages and takes a placating stance, and finally (4) the child returns to appropriate behavior. This chain reinforces the inappropriate behavior and increases the likelihood of such behaviors in the future (Eddy et al., 2001; Patterson et al., 1992). Over time, the characteristics of the parent-child interaction shape the child's working model of relationships. The child's working model comes to expect punishment, conflict, and rejection in relationships.

Classroom interactions. As these students enter school, this coercive cycle extends to teachers, decreasing the child's ability to benefit from positive educational and social opportunities (Burke et al., 2011). Furthermore, teachers are often not adequately trained in classroom management and inadvertently escalate aggressive and disruptive behaviors through coercive interactions (Kellam et al., 1998). In behavioral terms, coercive interactions occur when at least one of the members involved emit behaviors (stimuli) that are aversive to the other. When students emit disruptive behavior teachers find aversive, teachers are likely to engage in responses that will escape or terminate the aversive interaction, resulting in negative reinforcement of the teacher's behavior. Teachers who have students with behavior problems may notice coercive behaviors control interactions with these students, making exchanges difficult and unpleasant (Burke et al., 2011). Additionally, teacher reinforcement for positive behavior is often infrequent and reprimands given to problematic students are often non-contingent upon student behavior. Teachers who spend more time focusing on inappropriate behaviors than appropriate may maintain and even increase aggressive behaviors. The level of teacher's disciplinary actions is highly related to a student's level of rebellion and self-

reported delinquency, such that schools characterized by low achievement and high levels of antisocial behaviors tend to rely on suspensions and expulsions as the preferred response (Le Blanc, Vallieres, & McDuff, 1992; McEvoy & Welker, 2000). In schools with the worst discipline issues, rules are unclear, unfair, or inconsistently enforced; responses to student behavior are ambiguous or indirect; teachers do not know or disagree with the rules; teachers ignore misconduct; and students do not believe the legitimacy of rules (Burke et al., 2011).

Alternatively, schools and teachers can exert positive influences on students despite conditions in the home, social status, gender, race, or ethnicity (McEvoy & Welker, 2000). Implementing evidence-based classroom management practices may in the long-term decrease antisocial behavior in youth (Reinke & Herman, 2002). These practices include establishing and maintaining clear expectations for behavior, actively supervising student behavior, providing opportunities to respond, praising students for appropriate behavior, and giving error corrections for inappropriate behavior. Overall, considerable research has found that these strategies can reduce disruptive behavior (Kellam et al., 1998), and enhance academic achievement, school readiness, and students' social competence (Burke et al., 2011).

The limited use of classroom management techniques has been attributed to teacher knowledge, teacher's philosophical views, time demands, and the availability of professional development in classroom management (Noell & Gresham, 1993). However, teachers can reduce disruptive student behavior and increase academic engagement through the use of evidence-based classroom management practices (Simonsen,

Fairbanks, Briesch, Myers, & Sugai, 2008). In addition, in-depth training (i.e., through explicit instruction) and consultation in combination with performance feedback (Reinke, Lewis-Palmer, & Merrell, 2008; Simonsen, Myers, & DeLuca, 2010) may increase teachers' treatment integrity of these evidence-based classroom management strategies.

Evidence-Based Classroom Management Practices

As mentioned previously, evidence-based classroom management practices can reduce disruptive student behavior and increase academic engagement. In an effort to identify current evidence-based classroom management strategies, Simonsen and colleagues (2008) conducted a search of the empirical literature. Using criteria similar to those of What Works Clearinghouse (WWC; U.S. Department of Education, 2005), practices were considered evidence-based if they were (a) evaluated using sound experimental design and methodology (group experimental, group quasi-experimental, experimental single-subject designs, or causal comparative); (b) demonstrated to be effective; and (c) supported by at least three empirical studies published in peer-reviewed journals. The literature search yielded 20 general practices meeting evidence-based criteria. Of particular relevance to the present study are specific praise, error correction, prompts/pre-corrections, active supervision, and opportunities to respond (OTRs).

Specific Praise. Specific, contingent praise is one of the simplest and most empirically validated classroom management practices (Simonsen et al., 2008). Specific praise has consistently been found to decrease inappropriate behaviors and increase student engagement. Specific, contingent praise is a positive statement given by the

teacher when a desired behavior occurs acknowledging specifically what the student did well (Myers, Simonsen, & Sugai, 2011).

While praise is considered to have a substantial research base, some researchers argue that the use of extrinsic rewards (i.e., praise, tokens, and edibles) may decrease students' intrinsic motivation for the activity being rewarded (Deci, 1971, 1976; Lepper & Greene, 1975). This is known as the overjustification effect (Lepper & Greene, 1975). Intrinsic motivation is a behavior that is driven by internal rewards or is intrinsically rewarding. Alternatively, extrinsic rewards are external to the behavior. Some teachers are cited as viewing these rewards or reinforcers as "bribes" and fear students will become dependent on them rather than experiencing inherent motivation for learning. However, much of the research in this area suggests this concern can be ameliorated.

For a behavior to be considered intrinsically reinforcing, no external consequences should follow the behavior (Workman & Williams, 1980). If students were permitted to learn whatever content they desired, extrinsic rewards may not be necessary. Unfortunately, this arrangement is not feasible in classrooms and therefore some extrinsic rewards are required to engage students in important academic tasks and appropriate classroom behaviors. In addition, some students may exhibit performance deficits whereby they have the skills but due to various reasons such as academic frustration they do not find the academic environment as intrinsically motivating (VanDerHeyden & Witt, 2008). Similarly, some students may have a skill deficit that requires instruction in the desired behavior. Initially, continuous reinforcement (i.e., every time the behavior occurs) can be used to create a strong association between the behavior and the response

and increasing the rate of learning. Throughout the literature external rewards/reinforcement has been effective at increasing academic and behavioral outcomes (Workman & Williams, 1980).

One such study was a meta-analysis conducted by Cameron and Pierce (1994) examining the effects of reinforcement/reward on intrinsic motivation. Results indicated that external rewards do not decrease intrinsic motivation on any of the four included measures (free time on-task once reward is withdrawn, self-reports of attitude, performance during free-time measures, and willingness to volunteer for future studies without reward). When interaction effects were evaluated, findings indicated verbal praise actually produced an increase in intrinsic motivation. The only negative effect appeared to be when expected tangible rewards were given to students simply for doing a task.

Similarly, using an ABAB withdrawal design Sutherland and colleagues (2000) examined the effects of an observation-feedback intervention on the rate of teachers' behavior-specific praise and on-task behavior of students with emotional and behavioral disorders (EBD). Direct observation data were used to record behavior-specific praise statements and students' on-task behavior. The observation-feedback intervention consisted of the observer providing the teacher with verbal feedback on the observed rate of behavior-specific praise. Both behavior-specific praise and general praise was recorded. Results were consistent with previous research on the effect of teacher praise on the on-task behavior of students in general education classrooms. The percentage of

on-task behavior increased when the rate of behavior-specific praise increased and decreased when the rate of behavior specific praise was decreased.

In a more recent study, Myers, Simonsen, and Sugai (2011) used a multiple baseline design across teachers to evaluate the effects of a response-to-intervention approach on rates of behavior specific praise. Direct observation data were collected on the rate of specific behavior praise, general praise statements, rate of negative (i.e., corrections, reprimands) interactions with students, and the ratio of positive to negative interactions. Effects of the intervention on student academic, off-task, and disruptive behavior were also measured. Teachers received brief consultation consisting of rationale and examples of specific, contingent praise, data on teacher performance, and weekly praise from the researcher contingent on improved rates of specific, contingent praise statements. Similar to past research, performance feedback resulted in teachers' increased use of behavior praise statements. The data also demonstrated an overall downward trend in student problem behavior (i.e., off-task and disruptive behavior) in each classroom.

Error Correction. Error correction also represents a simple and evidence-based classroom management practice, which can decrease the likelihood of disruptive behavior (Simonsen et al., 2008). Brief, contingent, and specific error correction is an informative statement that is given when an undesired behavior occurs (contingent), states the observed behavior and tells the student exactly what they should do in the future in a brief, concise manner. Teacher use of brief, contingent, and specific error correction will also be targeted in the present study.

Brief, contingent, and specific error correction has been found to improve both academic and social behavior outcomes. In terms of academic outcomes, specific error corrections increased future success rates (i.e., decreasing errors) and improved word recognition and reading comprehension (Simonsen et al., 2008). Providing brief, contingent, and specific error correction also led to decreases in undesired behaviors. Error corrections that are loud in tone are less effective than quiet or discreet corrections. Similarly, error corrections that are brief (i.e., one to two words) are more effective than longer error corrections (i.e., two or more phrases), and corrections delivered consistently are superior to those delivered inconsistently.

More recently, Matheson and Shriver (2005) examined the effectiveness of command training with teachers on students' compliance rates and academic engagement. Teachers were taught how to provide effective commands through behavioral consultation. Effective commands were considered concise instructions that (a) elicit a distinct outcome, (b) are precise and temporally isolated, (c) are specific and direct, and (d) are given one at a time. Effective commands also included the qualities of a quiet voice tone, directive, stated positively, and descriptive. Ineffective commands were considered commands that did not include the characteristics of effective commands. During training, teachers were asked to verbally alter ineffective commands to effective commands and were provided performance feedback from the consultant on their success rate. Improvements in rates of compliance and academic behaviors were observed when teachers increased the use of effective commands. An improvement in compliance and academic behaviors were increased further when teachers used praise in

conjunction with more effective commands. Overall, results suggested teacher training in effective commands could increase teachers' use of effective commands and improve student outcomes.

Prompts/Pre-Corrections and Active Supervision. Pre-corrections are specific cues that provide students with information about the behavior desired in specific situations. (De Pry & Sugai, 2002). The pre-correction may be verbal (e.g., restatement of the desired behavior) or nonverbal (e.g., gesture). Active supervision is defined as the teacher moving, looking around, interacting with students, providing error correction (i.e., correcting behavior inconsistent with expectations), and delivering reinforcement (e.g., specific praise) for behavior consistent with expectations. Active supervision has been associated with decreases in off-task and disruptive behavior, as well as increases in academic engagement (Simonsen et al., 2008). Instruction in expectations coupled with reinforcement and error correction leads to the largest gains. Active supervision has also produced a classwide decrease in minor behavioral incidents (De Pry & Sugai, 2002) and higher levels of active participation (Schuldheisz & van der Mars, 2001).

Interestingly, in a study by Colvin and colleagues (1997), results indicated that the degree of active supervision, and not the supervisor-to-student ratio, accounted for the most variance in problem behavior. The authors were examining the effects of establishing expectations and active supervision on the transition behavior of students. Overall results indicated a substantial decrease in problem behavior for each transition setting (i.e., entering the school building, going to the cafeteria, and leaving the classroom at the end of the school day). Additionally, a correlation of -0.83 was found

between indices of problem behavior and number of interactions between supervising staff and students. This suggests that the higher the number of interactions between supervising staff and students, the fewer indices of problem behavior occur.

In a comparable study, Lewis and colleagues (2000) examined the effectiveness of similar procedures to the Colvin (1997) study on problem behaviors displayed during recess. Establishing expectations was part of a schoolwide social skills program, which consisted of identifying problems exhibited by the students at recess and identifying expected or replacement responses for the problem behaviors. Printed expectations for recess were reviewed and pre-correction regarding rules was utilized prior to students leaving for recess. Active supervision consisted of training playground monitors in the critical features of active supervision (i.e., move around, look around, and interact with students). Data were collected on the rate of student and playground monitor behavior. The intervention reduced overall rate of observed problem behavior during unstructured activities: however, an increase in active supervision on the part of the playground monitors was not observed. Reaction to the novelty and being observed may have contributed to a lack of observed differences in the monitors' behavior.

Within the classroom setting, De Pry and Sugai (2002) examined the effects of using active supervision, establishing expectations (i.e., pre-corrections) and daily data review on occurrences of minor behavioral incidences. As with the previously described studies, behavioral expectations were taught as part of a schoolwide implementation of behavioral support. Active supervision was considered as the teacher (a) circulating around the classroom, (b) scanning the classroom, (c) interacting with students, and (d)

reinforcing demonstrations of expected academic and social behaviors. Daily data review consisted of brief meetings with the teacher whereby the researcher reviewed graphed data of the teacher's performance. Minor behavioral incident data were collected throughout baseline and intervention phases. Teacher use of active supervision and/or pre-correction of behavioral expectations were also recorded. Results indicated a functional relationship between the use of active supervision and establishing expectations and subsequent decreases in minor behavioral incidents.

Opportunities to Respond. Another aspect of evidence-based classroom management practices targeted in the present study is opportunities to respond. An opportunity to respond is something a teacher does that elicits a student response (e.g., asking a question, presenting a demand). Common methods to increase the rate of presenting OTRs in the classroom are choral responding (i.e., students answering a question in unison) and response cards (i.e., erasable boards where students write their answers to a question and then hold up the boards for the teacher to see). Increasing the rate of opportunities to respond has been associated with positive effects on both student achievement and behavior.

For instance, Sutherland and Wehby (2001) conducted a review of the literature and examined the effects of increased OTR on academic and behavioral outcomes of students with emotional and behavioral disorders. The researchers suggested that the relationship between instruction and problem behavior could be used to reduce the academic difficulties of students with EBD and decrease the levels of disruptive behavior. Criteria for inclusion in the review included having participants with EBD or

students identified as exhibiting behavior characteristics of EBD, such as off-task, disruptive, or aggressive behavior. In addition, studies had to examine the effects of increased rates of OTR on one or more dependent variables of a behavioral or academic nature and have been published in a peer-reviewed journal. Overall, increased rates of OTR yielded improved academic outcomes, increased task engagement, and decreased disruptive student behavior. More specifically, reading and math outcomes were positively affected, students increased rates of correct responding, positive effects for task engagement were found, and instructional time in the classroom was used more efficiently.

Professional Development for Improving Classroom Management Practices

Despite the importance of classroom management, teachers typically receive little training in this area (Begeny & Martens, 2006). Furthermore, one of the most frequently cited reasons by teachers for leaving the profession or transferring schools is student disruptive behavior (USDOE, 2005). This is consistent with prior surveys, indicating student behavior problems as a primary concern for teachers and administrators (Elam, Rose, & Gallup, 1996). Evidence also suggests that teachers provide less instruction to students who exhibit problem behaviors (Carr, Taylor, & Robinson, 1991; Wehby, Symons, Canale, & Go, 1998). For instance, Carr and colleagues (1991) found that teachers provided more instruction to students exhibiting appropriate behavior than those exhibiting inappropriate behaviors. Fortunately, teacher implementation of evidence-based classroom management practices can be improved and supported through the use of school-based professional development activities such as coaching and consultation

(Reinke & Herman, 2002).

Traditional models of professional development that consist of a single meeting or information session have shown little effectiveness in influencing teacher behavior (Kinkead, 2007). This is due to their lack of ongoing contextual support to improve learning. One area of professional development that has shown potential at improving the impact of professional development efforts is decreasing professional isolation through interactive activities that connects teachers to other professionals such as consultants or coaches (McIntyre, Kyle, Chen, Munoz, & Beldon, 2010). As a result, researchers have examined the extent to which consultation and coaching have been effective in school settings.

Research on the effectiveness of consultation and coaching. Consultation and coaching models have demonstrated effectiveness at enhancing classroom practices, including emotional climate (Brown, Jones, LaRusso, & Aber, 2010), behavior management (Raver et al., 2008), and rules and routines (Rimm-Kaufman & Sawyer, 2004). Developers of these models suggest they work because they are responsive to teacher needs and strengths, provide supportive and specific feedback about practices, and offer coaching in research-based practices (Cappella et al., 2012). Furthermore, Driscoll and colleagues (2010) reported teachers were 13 times more likely to implement an intervention when given additional supports, such as a coach or consultant. When provided this support, teachers are also more effective intervention implementers and report greater self-efficacy and ability to maintain newly learned practices (Forman, Olin, Hoagwood, Crowe, & Saka, 2009).

While the interest in coaching is growing, the majority of available studies have focused on academic content coaching and curriculum implementation (e.g., reading, math, or science; Hershfeldt, Pell, Sechrest, Pas, & Bradshaw, 2012). Cappella and colleagues (2012) did find positive effects of the Bridging Mental Health and Education in Urban Schools (BRIDGE; Cappella et al., 2011) model, which sought to increase classroom interactions and address student behavioral challenges. Similar to other coaching models, BRIDGE promotes responsiveness to teachers, specific feedback, and coaching in context. However, BRIDGE also integrates a focus on universal and targeted support within an observational framework, and embeds these components within a sustainable delivery system. Mental health coaches implement the BRIDGE intervention with teachers in their classrooms. Those in the combined coaching and training condition experienced significant improvements in the closeness of teacher-student relationships, students' academic self-concept, and students' experience of victimization of peers relative to a condition of training alone.

In a similar study, the MyTeachingPartner program (MTP; Pianta, Mashburn, Downer, & Hamre, 2008a) utilized a video-based coaching model to enhance the quality of teacher-student interactions in classrooms. Video exemplars consisted of high-quality teacher-student interactions tied to specific dimensions of the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008b). The CLASS is an observational measure of teaching quality, which scores videotapes of teachers' practices. Coaching processes consisted of regular, multi-modal, ongoing performance feedback targeted to pre-k teachers through a standardized protocol (i.e., the CLASS). Scores are

used as a benchmark by the coach to identify target teacher behaviors for feedback and to set goals through collaborative discussion.

In a randomized trial of MyTeachingPartner at the pre-kindergarten level, one condition received only on-demand access to video-clip exemplars and the other the same access to video-clip exemplars as well as the web-mediated MTP coaching process (Pianta, Mashburn, Downer, & Hamre, 2008a). Teachers assigned to receive online coaching and feedback showed significant improvements on ratings of the quality of interactions. A second trial in secondary schools produced similar effects on student-teacher interactions and student achievement (Allen, Pianta, Gregory, Mikami, & Lun, 2011).

Recently, Becker and colleagues (2013) explored the relationship between coaching and implementation of the Good Behavior Game (GBG) with urban elementary school teachers. Researchers wanted to evaluate the nature, dosage, and sequencing of coaching activities. The GBG is a group-based token economy, where the groups or "teams" are reinforced for their collective success at reducing inappropriate behavior. Teachers attended a one-day GBG training, with support from the coaches. Coaches were expected to meet with each teacher approximately once a week and follow a two-phased coaching model (Becker, Bradshaw, Domitrovich, & Ialongo, 2013). In the universal coaching phase, which lasted approximately 4-6 weeks, coaches used the same strategies with all teachers. These activities included check-ins, modeling, needs assessments (e.g., observations), and technical assistance/performance feedback. In the second "tailored"

coaching phase, coaches developed individualized plans regarding the type and intensity of coaching support needed.

Findings were consistent with an adaptive model of coaching, whereby coaches strategically varied their coaching efforts based on teacher implementation quality. During the universal phase, teachers received coaching at the same frequency and duration regardless of implementation quality. During the tailored phase, coaches spent more time with teachers who demonstrated low implementation quality. Overall, coaching was associated with improved implementation quality of the GBG.

In a randomized controlled trial of schoolwide positive behavioral interventions and supports (SWPBIS), the PBISplus coaching model was evaluated (Bradshaw, Pas, Goldweber, Rosenberg, & Leaf, 2012). The study sought to support elementary classroom teachers in their acquisition and implementation of evidence-based classroom management practices as well as the implementation of Tier 2 targeted interventions. Coaches consulted with teachers, provided support in the use of evidence-based practices, and supported problem solving of student issues through a behavioral approach. The PBISplus coaching model integrated the technical, collaborative, and reflective coaching approaches to assist schools (Denton & Hasbrouck, 2009). Techniques included observations and providing feedback to teachers, using modeling of evidence-based tools, and delivering formal professional development sessions. Overall, school-level longitudinal analyses indicated that schools were able to implement SWPBIS with high fidelity and produced significant reductions in student suspensions and office discipline

referrals. Due to the group randomized controlled design, however, the data collected on student outcomes cannot be directly tied to coaches' activities.

While these studies suggest coaching adds value to professional development efforts, one challenge within the literature is a lack of consensus over the operational definition of coaching (Hershfeldt et al., 2012). In addition, there is concern pertaining to how the coaching is conducted. This includes who the coach works with, the techniques utilized, and the model applied. Currently, there is limited research specifying the core components of coaching, including (a) how coaches spend their time (or how they should), (b) the techniques used to improve teacher practices, (c) the model used (e.g., expert, peer, or collaborative), and (d) the training coaches need to be effective (Denton & Hasbrouck, 2009). Without a clear operational definition and use of a model with clear components, it is difficult to determine the effectiveness of coaching. Research on consultation processes, however, has generally indicated utilizing a process model is effective in changing teacher behavior (Erchul & Sheridan, 2008). Hence, consultation may be an area of research that addresses the weaknesses of coaching.

School-Based Consultation

Consultation can be an effective method of addressing learning and behavioral problems in school settings (Wilkinson, 2006). Typically consultation involves an indirect problem-solving process between a specialist (e.g., school psychologist) and one or more persons (e.g., teachers) to address client (e.g., student) concerns (Sheridan, Welch, & Orme, 1996). Consultation efforts are typically applied to a single referral problem. Key components include an indirect and problem solving emphasis, in a

collegial and voluntary nature, and a focus on process and outcome. Most researchers concede that consultation is a way of enhancing the effectiveness of evidence-based interventions (EBI; Frank & Kratochwill, 2008). Therefore, the traditional goals of consultation are to (1) identify appropriate evidence based interventions, and (2) to ensure the intervention is implemented as prescribed. There are various models of consultation including, behavioral, mental health, and organizational development. Within these models are variations such as "collaborative consultation," "instructional consultation," "process consultation," and "resource/consulting teacher."

Of these models, behavioral consultation (BC) is considered the most popular and empirically supported consultation model (Wilkinson, 2006). The BC model consists of four parts: (1) problem identification, (2) problem analysis, (3) treatment implementation, and (4) treatment evaluation (Kratochwill & Bergan, 1990). In the problem identification stage, the consultant and consultee operationally define the problem behavior in observable and measurable terms and form a hypothesis regarding triggering antecedents and maintaining consequences within the classroom. Then a plan is made to observe and measure the frequency and magnitude of the problem behavior. During the problem analysis phase the data is analyzed to either support or reject the functional hypothesis and design an intervention plan. Then during treatment implementation the consultee implements the designed intervention and the consultant optimizes the consultee's ability to implement the plan. This process includes teaching the consultee skills necessary for intervention, monitoring treatment integrity, and revising the intervention as needed. Lastly, in the treatment evaluation phase the consultant and consultee evaluate whether

the goal for the student has been met, whether the intervention was effective, and conduct post-implementation planning for maintenance and generalization.

Numerous studies have found behavioral consultation to be correlated with positive gains in student achievement and social behaviors (Sheridan et al., 1996). For instance, Noll and colleagues (1993) presented a case study of a pre-referral intervention for students with behavioral disorders using a behavioral consultation model. Four itinerant teachers provided consultation across 10 elementary schools. Services included (a) consultation and diagnostic support, (b) consultation and direct services for students identified as having behavioral disorders or learning disabilities, yet remained in regular classrooms with support, and (c) consultative services to teachers without direct services for students (i.e., defined as a minimum of three contacts).

Pre-referral intervention services were provided for 6 to 12 weeks and number of contacts consultants had with teachers ranged from three to five sessions per week, for 30 to 45 minutes. Interventions consisted of two types, behavioral interventions implemented by the general education teacher following consultation to increase appropriate and/or decrease inappropriate behaviors, and social skills programs implemented by the itinerant teacher. Results over a three-year span indicated 43% to 64% of the elementary school students served by itinerant teachers remained in general education classrooms without further support. Additionally, 14% to 22% of cases were identified as having other disabilities, and 23% to 39% were identified as having behavioral disorders, with half of those able to continue regular classroom placement. While this study provided initial support for the BC model, there were several limitations

worth noting. Most notably, the study was a case study indicating more rigorous methodology is necessary to evaluate BC's effectiveness. Additionally, the success of pre-referral interventions was attributed to the willingness of teachers to change and improve students' environments.

In another study performed by Fuchs and Fuchs (1989), three increasingly inclusive versions of the behavioral consultation model on problem behavior were assessed. Four experimental schools and five control schools participated, with 10 consultants supporting experimental schools. In experimental schools, teachers identified 48 of their most difficult-to-teach students at risk for special education referral or grade retention. Outcome measures included teacher ratings of target behaviors, remaining behaviors, and the combination of target and other behaviors as well as classroom observations.

Various components of the BC model were altered to determine the relative value of each stage of the model (Fuchs & Fuchs, 1989). In the least inclusive variation (Problem Identification and Problem Analysis; BC1) consultant and teacher worked together to identify and analyze the problem, however, the consultant neither assisted in nor monitored the teacher's implementation of the intervention. Additionally, intervention effects were not evaluated in a formative manner. The second variant of BC (Problem Identification, Problem Analysis, and Plan Implementation; BC2) also included the first two stages. In addition, BC2 required the consultant make a minimum of two classroom visits, whereby the consultant observed the teacher implement the intervention and provided corrective feedback. As with the first model, the BC2 model did not include

a formative evaluation stage. Lastly, the most inclusive version (BC3) included all three stages (Problem Identification, Problem Analysis, and Plan Implementation) as well as a formative evaluation stage.

Based on teacher ratings of students' problem behavior, BC2 and BC3 generated greater decreases than the control group (Fuchs & Fuchs, 1989). There was no such difference between students in the BC1 and control groups. There was also no difference in change of ratings for students in BC2 versus BC3, which may be due to four of eight teachers in BC3 not completing the consultative sequence. Overall, teacher perspective indicated that inclusive versions of BC were more effective than the least inclusive variant of the model. Observational data was somewhat consistent, whereby control students did not display a pre-to-post intervention decrease in target behaviors, but rather an increase (9%). Students in the BC1 model demonstrated a modest (8%) decrease in target behaviors. Surprisingly, BC2 students showed a similar magnitude of change (6%) as BC1 students, and BC3 students did not demonstrate improvement (0%). Researchers posit several possible explanations for the discrepant results. First, observational data may have failed to detect an improvement in the four 30-minute observations. Alternatively, observations, rather than ratings could be accurate and teachers participating in more inclusive consultation perceived a more positive transformation than teachers involved in the least inclusive version of BC. Lastly, researchers suggest both observations and ratings could be accurate but addressed different dimensions of behavior.

This study also had several limitations. For instance, rather than waiting for teachers to request help, consultants recruited teachers with difficult-to-teach students. This "proactive" consultant behavior questions whether recruited teachers are typical of those who normally participate in consultation. While this is a limitation and restricts generalization of results, these procedures align with typical recruiting procedures in the consultation literature. Another limitation is the lack of treatment fidelity data collected. Researchers only had general knowledge about the classroom-based interventions, but did not have information about the accuracy with which interventions were implemented.

More recently, MacLeod and colleagues (2001) examined the effectiveness of school-based behavioral consultation. Participants included 80 teachers who had partaken in consultation with school psychologists in the past 12 months. Consultation effectiveness was assessed through measures of consultant skills, quality indexes, and outcome indexes. Consultants' interpersonal, problem solving, consultation process, and ethical skills were rated as highly effective. Four of the six quality indexes (i.e., behavioral definition, direct measure/baseline, step-by-step plan, implemented as planned, results graphed, and results compared to baseline) were present in at least 68% of cases, suggesting the majority of critical elements of problem solving were included during consultation. Lastly, teachers reported that in two thirds of cases target student functioning improved and that the goals of consultation were achieved.

The author's highlight some limitations, such as the sample of teachers was small and it is possible only teachers with favorable impressions completed surveys. Additionally, the psychometric properties of the measures used are not well established,

indicating results should be interpreted with caution. Lastly, since a retrospective survey of consultees' perceptions was utilized it is unclear how accurately participants' recall of events depicts what occurred. Unfortunately, authors did not collect any observational data during the consultation process to determine teachers' level of adherence to the plan.

Overall, the research regarding behavioral consultation has yielded beneficial outcomes. However, as the above studies highlight, there are still some limitations within the literature. For instance, some of the studies did not utilize rigorous methodological designs (i.e., used case study designs). In addition, direct observational data of teacher use of targeted behaviors were not collected in some studies, reducing the generalizability of results. Lastly, some studies lacked treatment fidelity data, calling into question the change in student outcomes and whether it can accurately be attributed to the intervention (Wilkinson, 2006). Therefore, in order to address the previous gaps in the behavioral consultation literature, these limitations must be addressed in the present study.

Enhancing Treatment Integrity Through Performance Feedback

School-based consultation is generally accepted as a viable and acceptable means of service delivery in the schools (Sheridan et al., 1996). Despite the feasibility of consultation, research has indicated that a majority of teachers implemented interventions with low levels of treatment integrity within 7 to 10 days of initiating an intervention (Mortenson & Witt, 1998; Noell et al., 2000). Treatment integrity refers to the degree to which an intervention is implemented as designed (Gresham, 1989). Treatment integrity can be influenced by the strength of an intervention, intervention complexity, treatment agent competence, time required, resources, number of treatment agents, treatment agent

motivation, and treatment effectiveness. To make valid conclusions about intervention effectiveness, treatment integrity must be evaluated (Shadish, Cook, & Campbell, 2002). This relationship is most commonly conceptualized as a probabilistic one whereby as integrity decreases the probability of treatment failure increases (Noell, 2008). However, any decrease in integrity does not ensure a decrease in intervention effectiveness. This relationship is most notable when the deviation from treatment is small and a large effect is still observed. In addition, interventions that are poorly implemented are less likely to yield positive student outcomes (Noell et al., 2005) and change, or lack of change, in student outcomes cannot be attributed to an intervention if it is unknown whether it was implemented correctly (Wilkinson, 2006).

Currently there are several strategies to enhance treatment integrity, including performance feedback, constructing a graph either in the feedback session or prior with the teacher to increase consistency of effect, and training that includes enacting the treatment (Noell, 2008). Of these strategies, performance feedback is the one that has been found to be evidence-based (Fallen, Collier-Meek, Maggin, Sanetti, & Johnson, 2015). Performance feedback is a post-implementation strategy where a consultant presents a teacher with graphed student outcome data and treatment integrity data and reviews missed intervention steps. During the performance feedback process, the consultant observes the consultee, collecting objective data on the targeted student behavior and teacher treatment fidelity. At a later time the consultant and consultee meet to discuss the data. Positive teacher behaviors are praised and problem solving is conducted to enhance treatment fidelity. The process continues until the teacher reaches

the desired level of treatment fidelity. Numerous studies on the effectiveness of performance feedback have resulted in increased treatment integrity (Mortenson & Witt, 1998; Noell, Witt, Gilbertson, Rainer, & Freeland, 1997, 2000, 2005; Witt, Noell, LaFleur, & Monenson, 1997).

One such study conducted by Witt and colleagues (1997) provided performance feedback to four general education teachers implementing an academic performance intervention. Training consisted of didactic training with consultant and teacher, student training with consultant supervision, and in-vivo training on the first treatment day to ensure accurate implementation. After training, teachers implemented the intervention independently and baseline treatment integrity data were collected from permanent products. Concluding baseline, daily performance feedback was conducted. Finally, a maintenance phase was applied where the teacher again implemented the intervention independently.

On the training day, teachers implemented the intervention with 100% fidelity, however, during baseline, integrity began to trend downward (Witt et al., 1997). During the performance feedback phase implementation trends increased and maintained high levels. This pattern continued for three of the four teachers into the maintenance phase. Overall, results provided initial evidence for the influence of performance feedback to increase intervention integrity, as well as to maintain integrity after feedback was removed. However, one limitation is the rigorous training methods used during the training period may have enhanced the teachers' later implementation fidelity.

To determine if results would be similar when the training phase was less

intensive and more aligned with practical consultation practices in schools Noell and colleagues (1997) conducted a replication of the Witt et al. (1997) study. The procedures were the same as the Witt study, however, training was conducted in a more traditional consultation format with the consultant explaining how to implement the intervention during a problem analysis meeting. Then, teachers independently implemented the intervention and had no further communication with the consultant until the performance feedback phase. Performance feedback was also conducted similarly to the Witt et al. (1997) study.

Findings were similar to those found in the Witt et al. (1997) study in that teachers demonstrated high levels of fidelity for 2 to 3 days after training and then a declining trend during baseline (Noell et al., 1997). During the performance feedback phase, all of the teachers' treatment fidelity increased, which continued into the maintenance phase. Both of these studies provide evidence for performance feedback as a tool to increase intervention implementation integrity.

Since these studies, other researchers have examined procedural elements of performance feedback that may be adapted within consultation. For instance, Mortensen and Witt (1998) examined whether the frequency of feedback meetings influenced treatment fidelity of a reinforcement based intervention for academic performance deficits. Again, procedures were similar to Witt et al., (1997); however, performance feedback sessions were weekly rather than daily. Treatment integrity decreased during the independent baseline implementation phase. During the weekly performance feedback phase, treatment integrity increased and effects continued into the maintenance

phase for two of the three teachers. Results indicated weekly performance feedback was effective for improving integrity. However, comparatively the effects were not as high as in the Witt et al. (1997) study. Overall, higher frequency meetings may result in a greater effect, but weekly feedback can still produce a desired effect.

Similarly, Noell and colleagues (2000) examined the effect of brief feedback meetings with and without the use of student and teacher graphs. Brief meetings without graphed data were effective for two of five student-teacher dyads. The subsequent phases of feedback with graphed data were effective for four of five. In a follow-up study, Noell and colleagues (2002) used comparable procedures and found similar results. Brief meetings without data were beneficial for one of four student/teacher dyads but meetings with data driven feedback improved fidelity for all four. While these studies provide some support for the inclusion of data review in the feedback process, results are somewhat contradictory and warrant further research.

Due to the increased popularity of performance feedback, Solomon and colleagues (2012) conducted a meta-analysis of single-case studies using performance feedback to increase treatment integrity. The authors were specifically interested in how effective performance feedback was for different student age ranges, special and general education teachers, and different types of interventions. In addition, they examined whether a delay of feedback affected the power of the intervention. After initial screening, 36 studies were included in analyses.

Overall, performance feedback resulted in significant behavioral change regardless of setting, dependent variable, delay of feedback, or type of intervention

(Solomon, Klein, & Politylo, 2012). The average effect across studies was moderate in size. Performance feedback was effective in preschool through high school and grade did not significantly moderate the effectiveness of performance feedback. The difference between general and special education was significant, suggesting performance feedback may be more effective for general education teachers. However, practical differences in effect sizes were small and performance feedback was effective for both types of educators. Performance feedback was more effective for academic interventions than behavioral, however, measurement of the construct of treatment integrity may have confounded results. Immediate feedback had higher effects than weekly feedback, although the difference was not significant. Some moderators were unable to be analyzed, including the use of graphs during performance feedback. Based on the findings, the authors suggest further research investigating performance feedback to target the treatment integrity of behavioral interventions, the delay of performance feedback, and the additive benefits of different components such as graphs are warranted.

More recently, Fallon and colleagues (2015) conducted a systematic review of performance feedback following What Works Clearinghouse guidelines (Kratochwill et al., 2010). Results indicated 47.9% of studies met design standards and 26.6% met design standards with reservation. Performance feedback was most often provided during individual, in-person meetings and using verbal feedback. Most studies included graphic performance feedback or problem solving around implementation issues. Almost half of studies utilized all three components into performance, suggesting these elements may be part of a standardized performance feedback protocol. In general, strong to moderate

evidence was provided to determine that performance feedback could be considered an evidence-based practice.

Lately, researches have applied performance feedback to increasing teachers' use of evidence-based classroom management practices. Using performance feedback to address classroom management practices rather than focusing on individual students' behavior may be more efficient in that it reduces current student behavioral difficulties at a broader scale (Reinke et al., 2008). However, there is limited research in the area of performance feedback with a teacher to increase classroom management practices. Additionally, many of the existing studies are plagued by limitations and lack replications, restricting their application to school settings.

Previous research on performance feedback and classroom management. Rienke and colleagues (2008) conducted one such study where the effects of performance feedback on teacher use of specific praise were evaluated. The authors developed the Classroom Check-Up (CCU; Reinke et al., 2008) model to address teacher use of praise. The CCU is grounded in motivational interviewing, including using personalized feedback to teachers on classroom behaviors, encouraging personal responsibility for decision making, development of a menu of options for interventions, and supporting teacher self-efficacy. This is completed through a series of steps: (1) assessing the classroom, (2) providing teachers with feedback, (3) developing a menu of interventions, (4) choosing the intervention collaboratively with the teacher, and (5) having the teacher self-monitor implementation of the intervention. Following the CCU, teachers are provided daily visual performance feedback. Visual performance feedback consisted of

teachers receiving a line graph depicting the rate of specific praise and classroom disruptive behaviors observed in the classroom. Data were not discussed with the teacher.

Overall, rates of praise did not consistently increase following CCU/self-monitoring, however, praise increased for all teachers once performance feedback was implemented. During baseline, all classrooms had higher rates of classroom disruptions than praise. In the CCU/self-monitoring phase, two of four classrooms demonstrated an initial downward trend in disruptive behavior and a slight increase in praise (Reinke et al., 2008). One classroom demonstrated no noticeable change. During visual performance feedback, three out of four classroom's rate of praise was higher than rate of disruptions. Classroom three had no data points during performance feedback, in which the rate of praise was lower than the rate of disruptions. Classroom four had low rates of classroom disruptions and praise during baseline. During CCU/self-monitoring and visual performance feedback phases, classroom four increased the rate of praise, but had little change in the rate of disruptions. At a one-month follow-up, all four classrooms indicated higher rates of praise and lower rates of disruption, but downward trends in praise were observed across classrooms.

This study had several limitations. For instance, there was no direct observation of treatment integrity (Reinke et al., 2008). Treatment integrity data were only collected via teacher self-report. Some of the inter-rater agreement data were also below the 80% threshold (Kratochwill et al., 2010). In addition, performance feedback was provided through visual means only. Successful performance feedback interventions typically include a review of data on teacher performance, praise for correct implementation,

corrective feedback on procedures used incorrectly or infrequently, problem solving, and addressing questions (Noell et al., 2000). Performance feedback in this study consisted of the consultant handing the teacher a graph once a day, without discussion. Despite the variation in performance feedback components, the researchers suggest that CCU/self-monitoring plus visual performance increased rates of praise for all teachers, however, classroom three had no data points during performance feedback, in which the rate of praise was lower than the rate of disruptions. Furthermore, classroom three yielded the second highest effect size for the CCU plus visual performance feedback phase according to Cohen's (1988) guidelines. If classroom three had no data points during performance feedback, it is unclear how an effect size was calculated and subsequently of one of the highest magnitudes.

In addition, researchers calculated the standard mean difference for each variable in the CCU/self-monitoring phase and the CCU/self-monitoring plus visual performance feedback phase to determine effect sizes. While this method can be easy to use, it is insensitive to data trends (e.g., positive or negative slope), which is a key feature of single case design (Maggin et al., 2011). This approach also does not account for the dependence between observations (i.e., autocorrelation). Such dependence violates the parametric requirement that residual errors not be correlated and subsequently leads to either inflated or deflated standard errors depending on the direction of the autocorrelation. The presence of even small and non-significant autocorrelations can increase the Type I or II error rate.

Similarly, MacSuga and Simonsen's (2011) explored the effects of performance feedback with teachers and the use of classroom management practices. The model utilized included (a) a classroom management checklist that teachers can use to self-assess across time and (b) consultation incorporating action planning and performance feedback. The process is divided into four phases: (a) initiation, (b) promoting skill acquisition, (c) building skill fluency, and (d) supporting skill maintenance.

Results of two case studies indicated teachers increased use of evidence-based strategies when the model was implemented (MacSuga & Simonsen, 2011). During baseline, one teacher's use of strategies was fairly low and stable and students demonstrated low rates of on-task behavior. The other teacher implemented half of the strategies either partially or fully prior to consultation. Following the first consultation meeting, the first teacher demonstrated minimal progress and her students demonstrated even less on-task behavior. At the midpoint consultation meeting, daily performance feedback was initiated and overall implementation increased and students engaged in more on-task behavior. The other teacher reached nearly full implementation of strategies with only one consultation meeting and student on-task behavior increased.

While results were promising, there are several limitations associated with this study. A major limitation is the model was piloted using a descriptive single case study (AB) design with a baseline (A) phase before consultation and an intervention (B) phase during consultation. An AB design does not clearly demonstrate a functional relationship between the independent and dependent variables (Kennedy, 2005), which is a threat to internal validity. Furthermore, without additional phases, there is no control for other

events occurring during treatment (i.e., history or maturation effects). More rigorous experimental single case or group designs are necessary to examine the effects of using this consultation model with additional participants.

In addition, no inter-rater agreement data were collected, limiting the reliability of results. There was also no direct observation data collected on teacher use of classroom management strategies, which is a suggested component of the model and would be a valuable contribution during performance feedback. Similarly, students' were marked as on-task or off-task at the end of one-minute intervals throughout a 15-minute observation. A direct assessment measure designed to assess student academic behavior in the classroom environment using momentary time sampling may more accurately capture levels of on- and off-task student behavior. The model was not followed as outlined and also only included three consultation meetings, yielding inconsistent results in the consultation only phase. Lastly, a revised version of the checklist used has been created, the Classroom Management Self-Assessment (Simonsen, Fairbanks, Briesch, & Sugai, 2006), indicating further exploration of the model with the new checklist.

Lastly, Simonsen and colleagues (2010) examined the effects of explicit training and performance feedback on teachers' implementation of three classroom management skills (e.g., prompts for social behavior, academic opportunities to respond, and specific praise). The prompt-occasion-reinforce training (PORT; Simonsen, Myers, & DeLuca, 2010) intervention is composed of two phases, explicit training and performance feedback. During explicit training, discussion, activities, and self-management strategies

to promote generalized behavior change are used. In the second phase, performance feedback (e.g., data review, contingent praise, and error correction) is conducted.

Across the three teachers, baseline data were stable for both prompts and praise statements, there were no observed prompts for social behavior and low levels of specific praise (Simonsen et al., 2010). There was variability among all teachers during baseline for OTRs. After the introduction of training, there was an increase in implementation trend for prompts and specific praise. However, no clear change was observed for OTRs. During the performance feedback phase, there was an increase in implementation trend across all three behaviors. Overall, there was not a functional relationship between explicit training and teachers' use of classroom management skills. However, the introduction of performance feedback following training was functionally related to an increase in the level, trend, and stability of teachers' use of each skill.

While this study provides some support for performance feedback with teachers to increase teacher use of classroom management practices, there are several limitations worth noting. First, this study was conducted in an alternative setting and participating teachers each had more than 10 years of experience in regular education, special education in inclusive settings, special education in alternative settings, or some combination thereof. Therefore, participants may not represent "typical" teachers, limiting generalization of study results beyond the study sample. Due to the small sample size, researchers also only used visual analyses. Combining effect size analysis with visual analysis one can visually assess and confirm if effect sizes are an accurate depiction of the data and exclude potential confounding influences (Parker, Hagan-Burke,

& Vannest, 2007). The performance feedback provided was also intensive. Research suggests that weekly feedback may be equally effective (Mortensen & Witt, 1998).

In addition, while there were three participants, two of the feedback phases for prompts had less than five data points (e.g., three and four). When utilizing a multiple baseline design, there must be at least three attempts to demonstrate an intervention effect at three different time points or within three different phase repetitions to *Meet Standards* (Kratochwill et al., 2010). Furthermore, a minimum of six phases and at least five data points in each is necessary. The present study fails to replicate the intervention effect for prompts according to WWC standards due to fewer than five data points in two of the feedback phases. However, the study does have at least three data points per feedback phase, which could *Meet Standards with Reservations*.

Researchers also did not collect treatment integrity data on the PORT intervention, which is one of Horner and colleagues (2005) quality indicators of single case design. Without treatment integrity data, valid conclusions about intervention effectiveness cannot be made (Shadish et al., 2002). Similarly, inter-rater agreement data were collected on only 15% of observations. According to Kratochwill and colleagues (2010), inter-rater agreement data should be collected in each phase or on at least 20% of the data points in each condition. Inter-rater agreement did meet the .80 to .90 threshold. Lastly, student behavior was not directly measured. While previous research has found the effectiveness of increasing prompts (e.g., De Pry & Sugai, 2002), OTRs (e.g., Sutherland, Alder, & Gunter, 2003), and specific praise (e.g., Sutherland and Wehby, 2001) on increasing appropriate (and decreasing inappropriate) student behavior,

inferences regarding changes in teacher behavior affecting student behavior cannot be made.

While these studies attempt to address the lack of performance feedback and classroom management research, each is plagued with limitations. For instance, either the studies did not collect treatment integrity data or only used teacher self-report, which can often be inflated. Similarly, one of the studies reported inadequate inter-rater agreement and another failed to even calculate agreement. The performance feedback procedures also varied within the models, with all providing daily feedback using either a checklist, email and verbal feedback, or through visual means. Outcome data were also varied, with some studies including student and teacher outcome data and others only using one outcome variable. Furthermore, none of the described study's methodology would meet WWC *Meets Evidence Standards*. Lastly, none of the proposed models have undergone replications, which enhances external validity.

Limitations of Past Research

Despite the evidence-base for performance feedback, many of these findings have not been applied to increasing the integrity of classroom management practices. For instance, Noell and colleagues (1997) found that a less rigorous training phase could be used prior to implementation and treatment integrity still improved. The studies pertaining to performance feedback and classroom management each had varying training methods that do not align with traditional consultation practices (Noell et al., 1997). Similarly, Mortensen and Witt (1998) suggest that, while effects may not be as high, weekly rather than daily performance feedback meetings can be effective. The

meta-analysis conducted by Solomon and colleagues (2012) yielded similar results. Weekly meetings also represent a less resource intensive option for consultants working in school settings. However, studies examining performance feedback to increase teacher use of classroom management practices have only consisted of daily performance feedback meetings, suggesting further research into the use of weekly meetings instead. Additionally, results regarding the use of student and teacher graphs are still somewhat inconclusive (Noell et al., 2000; Noell et al., 2002; Solomon et al., 2012) and not all studies targeting classroom management practices follow prescribed performance feedback procedures. Some studies used visual methods (Rienke et al., 2008), one used an author created checklist that has not been validated (MacSuga & Simonsen, 2011), and another provided verbal or email feedback based on teacher preference (Simonsen et al., 2010). These performance feedback procedures deviate from those conducted in foundational performance feedback studies (Mortenson & Witt, 1998; Noell et al., 1997; 2000; 2002; Witt et al., 1997) and these methods have not been validated. Additionally, Fallon and colleagues (2015) suggested verbal feedback might be an element of a standardized performance feedback protocol. Therefore, subsequent research addressing performance feedback and classroom management should examine the use of student and teacher graphs while following a standardized performance feedback protocol (Fallon et al., 2015; Noell et al., 2005). Lastly, the purpose of single case designs is to establish if there is a functional relationship between the independent and dependent variable, however, these findings are limited if rigorous methodology is not utilized. Consequently, sound methodology is preferred when making determinations about change in outcome

variables. While the previously discussed studies demonstrated positive results, they often lacked treatment integrity, adequate inter-rater agreement, and teacher and student outcome data, which limits generalizability. Furthermore, most studies would not meet What Work Clearinghouse's *Meets Evidence Standards* for single case designs (Kratochwill et al., 2010).

Rationale and Purpose of Present Study

Challenging student behaviors are impacting teachers' ability to teach (Markow et al., 2006). When teachers have to focus on disruptive classroom behavior, the amount of time devoted to instruction decreases (Reinke et al., 2007). In addition, disruptive student behavior is tied to long-term negative academic, behavioral, and social outcomes (Kellam et al., 1998; Reinke & Herman, 2002). While many schools are turning to suspension to address these behaviors, a better alternative to reduce disruptive student behavior and increase academic engagement is the use of evidence-based classroom management practices (Simonsen et al., 2008). Unfortunately, traditional models of professional development are often unsuccessful at changing teacher behavior (Fixsen et al., 2005; Kinkead, 2007). However the use of school-based consultation in combination with performance feedback can increase teachers' use of evidence-based classroom management practices (Reinke et al., 2008; Simonsen et al., 2010), which in turn can increase student engagement.

In general, performance feedback is an evidence-based practice that has been beneficial at enhancing the treatment integrity of behavioral intervention outcomes (Fallon et al., 2015; Solomon et al., 2012). However, few researchers have attempted to

use performance feedback to increase teachers' use of classroom management practices and existing studies have limitations and lack replications. Therefore, a necessary extension of the research would be to examine the relationship between performance feedback with a teacher and teacher implementation fidelity of classroom management practices and student outcomes. The proposed study will advance the literature by attempting to address still unanswered questions in the performance feedback area. This will be achieved by utilizing training that aligns with traditional consultation, following performance feedback as previously described (i.e., Noell et al., 2005), and providing weekly performance feedback using graphs of teacher and student data. Furthermore, the proposed study will address previous limitations by using methodology that meets WWC *Meets Evidence Standards* and collects adequate treatment and inter-rater agreement data. This study will demonstrate the utility of performance feedback in the school setting. More importantly, this research will determine whether performance feedback with a teacher can decrease disruptive student behavior and increase student academic engagement as well as increase teachers' treatment integrity of evidence-based classroom management practices. Lastly, this study will identify if teachers find performance feedback procedures as socially valid (See Figure 1).

Thus, the purpose of the present study is to examine the relationship between performance feedback with a teacher and teacher and student behavioral outcomes. The following research questions will be examined:

1. To what extent is there a functional relationship between performance feedback with a teacher and student engagement?

2. To what extent is there a functional relationship between performance feedback with a teacher and disruptive student behavior?
3. To what extent is there a functional relationship between performance feedback with a teacher and teachers' treatment integrity of targeted classroom management practices (i.e., specific praise, error correction, prompts/pre-corrections, active supervision, and providing OTRs)?
4. To what extent do teachers find performance feedback as addressing socially significant goals, socially acceptable procedures, and socially important outcomes?

Methods

Setting and Participants

Participants were selected from an urban school district in Southern California. Demographic data from the 2013-2014 school year was Hispanic or Latino, 59.9%, White, not Hispanic, 24.6%, African American, 7.2%, Asian, 3.3%, Filipino, 1.2%, Pacific Islander and American Indian or Alaska Native, less than 1% (CDE, 2014). English Language Learners represented 17.3% of students. The percentage of students receiving free and reduced-price lunch was 64.7%.

School one's demographic data included Hispanic or Latino, 76.7%, African American, 8.3%, White, not Hispanic, 7.9%, Asian, 1.8%, Pacific Islander, 1.2%, Filipino and American Indian or Alaska Native, less than 1% (CDE, 2014). English Language Learners represented 37.9% of students. The percentage of students receiving free and reduced-price lunch was 90.3%. For school two, Hispanic or Latino was 65.2%,

White, not Hispanic, 22.1%, African American, 6.1%, Asian, 3.6%, Filipino, Pacific Islander, and American Indian or Alaska Native, less than 1% (CDE, 2014). English Language Learners represented 21.9% of students. The percentage of students receiving free and reduced-price lunch was 67.5%.

Teachers. Teacher participants were drawn from all available general education teachers in the above schools. Teachers were asked to volunteer in the consultation process and the order in which they participated in consultation was randomly determined through computer programming (i.e., Excel spreadsheet). During the Problem Identification Interview, teachers were asked when they experience the most challenges with classroom management (i.e., certain instructional activities, transitions, time of day; Kratochwill & Bergan, 1990). Based on their response, an observational time was established and maintained throughout the 10 weeks of the study. Teacher demographic data, including age, ethnicity, grade level, years of experience, amount of training in behavior management, level of education was collected using a survey (See Figure 2). Questions regarding previous classroom management training and perceptions of administrative support in addressing behavior concerns were also included in the survey. In addition, classroom demographic data (i.e., number of students, gender, number of English language learners, number of office discipline referrals) was collected. See Table 3 for a summary of teacher and classroom demographic data.

Measures

Classroom management and student engagement. Classroom management and student engagement data were collected through direct observation procedures utilizing

the Classroom Observation of Teachers and Students (COTS) measure created by the primary researcher (See Figure 3). Development of the measure followed the structure of the Behavioral Observation of Students in Schools (BOSS; Shapiro, 2004). The BOSS measures levels of on- and off-task student behavior. Reports of inter-observer agreement for the BOSS have been consistently high. Ota and DuPaul (2002) reported total agreement ranging between 90% and 100%. More recently, DuPaul et al. (2004) reported kappas ranging from .93 to .98 for observations in a large sample of children with ADHD and normal comparison children (N = 136).

Following recommendations within the BOSS manual (Shapiro, 2004), observation data will be collected for 20-minutes during the selected class period three times a week. The measure consists of 30-second intervals subdivided into two 15-second intervals. The first half of the 30-second interval measures targeted classroom management practices and the second half measures student engagement behaviors. Every 15-seconds alternates between classroom management and engagement variables with both recorded as one 30-second interval.

Classroom management practices. The observer recorded whether the teacher engaged in each of the targeted classroom management practices at any point during the first half of the 30-second interval. Targeted classroom management practices included specific praise, error corrections, prompts/pre-corrections, active supervision, and providing OTRs. Specific, contingent praise is a positive statement, typically provided by the teacher, when a desired behavior occurs (contingent) to inform students specifically what they did well. Brief, contingent, and specific error correction is an informative

statement that is given when an undesired behavior occurs (contingent), states the observed behavior, and tells the student exactly what behavior they should do in the future in a brief and concise manner. Prompts/pre-corrections are specific cues that provide students with information about the behavior desired in specific situations. Prompts may be verbal, nonverbal, or both. Active supervision is defined as the teacher moving, looking around, interacting with students, correcting errors made by students (i.e., behavior inconsistent with expectations), and providing reinforcement for behavior consistent with expectations. Opportunities to respond are teacher behaviors that prompt or solicit a student response (i.e., asking a question, presenting a demand). Examples include choral responding (i.e., students answering a question in unison) and response cards (i.e., erasable boards). Objective definitions, examples, and non-examples of these five practices can be seen in Table 1.

Student engagement. Student engagement was coded as either active engaged time (AET) or passive engaged time (PET; See Figure 3). According to Rathvon (2008), research has indicated that an effectively managed classroom has a rate of 80% academic engaged time or higher. Active engaged time is defined as when the student is actively attending to assigned work (i.e., writing, reading aloud, raising hand, talking to a peer about assigned material). A student will not be considered actively engaged if the student was talking to a peer, writing, or generally talking about something unrelated to class content. Passive engagement is when a student is passively attending to assigned work (i.e., listening to a lecture, looking at an academic worksheet, silently reading assigned material, listening to a peer respond to a question). A student is not considered passively

engaged if reading material not related to course content or looking at other objects in the room other than those that are part of instruction. Off-task behavior or non-engagement is defined as any activity not associated with classroom instruction. These include motor behaviors (i.e., out-of-seat, manipulating objects not related to academic task), verbal behaviors (i.e., whistling, humming, talking to another student about unrelated topic), and passive behaviors (i.e., looking around the room, staring out the window).

Following procedures used in previous research to obtain classwide student engagement (McKissick, Hawkins, Lentz, Hailley, & McGuire, 2010), the present study utilized a modified approach to measure student engagement. At the start of the second half of each 15-second interval, a new student was observed. The first student designated for observation was determined based on seating chart records. After each interval, the observer moved to the next student until all students were observed. This sequence continued across the observation period. Reports of average inter-rater agreement for the use of this procedure were 95% (range 91%-100%) across observation sessions.

Although the proposed direct observation measure was adapted from the BOSS, which has adequate reliability and validity, pilot testing was conducted to ensure the observational coding scheme described applies to participants in this study. A sample of five general education and five special education teachers were recruited from the previously described school district and trained observers conducted in-class observations. The selection of both general and special education teachers enhances generalization of the measure to different classroom formats. Inter-rater agreement was calculated using Cohen's kappa coefficient to determine consistency among raters

(Hintze, 2005). The inter-rater reliabilities for raters was .99 and ranged from .94-1.00. Content validity was also established by having advanced graduate students not involved in development determine whether the measure adequately sampled the domain of interest (Raykov & Marcoulides, 2011).

Observations took place in time sampling interval recording format. Observers wore headphones and listened to a recording where a bell rang every 15-seconds indicating the start of the interval. Interval recording data yields a percentage of the observed time that a behavior occurs. Because the behaviors observed vary, two types of interval recording were utilized. Active and passive engagement used momentary time sampling procedures. Within momentary time sampling, the observer records at the bell if a student is actively or passively on task at that moment. This is because these behaviors do not have a discrete start and stop period and have the potential to occur at a steady state (Hintze, Volpe, & Shapiro, 2008).

Classroom management practices (i.e., specific praise, error corrections, prompts/pre-corrections, active supervision, and providing OTRs) and student off-task behaviors used partial interval recording. Similar to momentary time sampling, these behaviors do not have a clear start and end, however, because they have more of an inconsistent duration and occur at a relatively smaller rate they are better measured through partial interval recording (Hintze et al., 2008). During partial interval recording, one or more behaviors are recorded if they occur during any portion of the observed interval. For instance, if specific praise took place during an interval where classroom management is being recorded then that instance is coded. However, if specific praise

occurs again during that interval, the instance is recorded only once.

Inter-rater reliability. To ensure reliability of direct observation data inter-rater agreement was collected in all phases and on at least 20% of all sessions (total across phases) for a condition (Kratochwill et al., 2010). Inter-rater agreement was calculated using Cohen's kappa coefficient because it corrects for chance between observers (Hintze, 2005). Hartmann, Barrios, and Wood (2004) suggest a minimum acceptable value of at least .60 if measured by Cohen's kappa. Inter-rater reliability can be improved by having raters work independently, informing raters their ratings will be checked, and most effectively providing training that emphasizes the distinction between observation and interpretation. Therefore, raters participated in a training consisting of (a) one meeting to introduce the coding forms and discuss operational definitions of the behaviors included on the forms and (b) two or more sessions of in-vivo training (i.e., observing teachers in the classroom) with the forms. Inter-rater reliabilities for raters was .99 and ranged from .95-1.00.

Consultation procedural integrity. Noell and colleagues (2005) defined treatment plan implementation (i.e., treatment integrity) as the degree to which a treatment plan developed within consultation is implemented as designed. In addition, treatment plan implementation is the outcome of the consultation process (e.g., dependent variable; DV) and is the most immediate and direct outcome of consultation. Alternatively, consultation procedural integrity (CPI) is the degree to which consultation procedures were implemented as designed in both practice and research contexts (i.e., independent variable; IV). Peterson and colleagues (1982) have indicated the emergence

of a “curious double standard” within research, whereby more rigorous standards have been applied to the operational definition and measurement of dependent variables than independent variables. In addition, Gresham, Gansle, & Noell (1993) found only 15% of behavior studies from 1980-1990 assessed the treatment integrity of IV implementation. Therefore, CPI was assessed in the present study through several methods. First, the consultant completed a fidelity checklist after consultation sessions to determine if the consultation procedures were implemented as designed (See Figure 4). Items pertained to the problem identification, problem analysis, weekly performance feedback, and the treatment evaluation meetings. Items included the following: “Identifies the problem in operational terms,” “Antecedents, consequences, and patterns are discussed,” "Receives graphed teacher and student behavioral data,” and “Evaluate goal attainment.” Second, each consultation session was videotaped (i.e., contingent on teacher consent). These videotapes were reviewed by an independent observer and coded for fidelity of implementation using the same checklist. Finally, inter-rater agreement was conducted and calculated for the evaluation of treatment integrity by the independent observer and consultant. Reliability was established for at least 20% of the consultation sessions. The interrater reliabilities for raters were 1.00 and procedural integrity was 100%.

Social validity. Social validity is the extent to which consumers view a given practice as addressing socially significant goals, socially acceptable treatment procedures, and socially important intervention outcomes (Wolf, 1978). A questionnaire was given to participating teachers to assess social validity from the teacher’s perspective of performance feedback during the final consultation meeting (See Figure 5). The

questionnaire was adapted from the Intervention Rating Profile-15 (IRP-15; Witt & Elliott, 1985) to measure teacher perceptions of the social significance of intervention goals, social acceptability of intervention procedures, and likelihood of socially important outcomes. The wording of the IRP-15 was modified slightly to reflect performance feedback as the intervention. The IRP-15 is a 15-item, factor-analytically derived survey whereby each item is rated on a 6-point Likert-type scale ranging from 1 (strongly disagree) to 6 (strongly agree). Internal consistencies range from .88 to .98. Higher scores on the questionnaire suggest higher acceptability.

Procedure

Upon approval from the university's Institutional Review Board and the participating school district, the primary researcher contacted principals for approval. Then teachers were asked to voluntarily participate and those who agreed were asked for written consent, including consent to be videotaped. Incentives were provided in the form of $25 gift certificates. Once each teacher was identified, the consultant and teacher met briefly to discuss the process. After the brief meetings, the primary researcher and trained graduate students collected baseline teacher and student observation data three times a week. Following the baseline phase, classroom observations continued three times a week. Once a week the teacher met for a performance feedback session with the consultant (i.e., primary researcher). This process continued for 10 weeks ending at the 10-week mark or until a stable level and trend of classroom management practices was observed for each teacher.

Baseline. Initial baselines for teacher behaviors provide information about the extent to which each teacher implements classroom management practices examined in this study (i.e., prompts/pre-corrections, active supervision, providing OTRs, specific praise, and error correction) independently. For each of the participating classrooms, collection of baseline teacher and student observation data began the week after initial meetings had been completed. When treatment integrity data were low and stable or trending downward, performance feedback meetings were initiated.

To ensure changes in teacher behavior are associated with the introduction of performance feedback and not the presence of a consultant, weekly meetings occurred both in baseline and in the treatment phase. This was due to meetings with consultants possibly providing teachers a prompt to implement the treatment (Noell et al., 2000). However, research examining this relationship has found inconsistent results. As mentioned previously, Noell and colleagues (2000; 2002) examined variations in follow-up meetings to determine whether the presence of the consultant versus providing performance feedback improved fidelity. Across studies, performance feedback led to the most improvements in implementation.

Weekly meetings in the present study's baseline were 5-10 minutes in length and addressed scheduling of direct observations and discussion of study procedures. Additionally, the teacher survey and Problem Identification Interview were administered (See Figures 2 and 6). Performance feedback on current classroom practices was not provided.

Consultation process. The consultant utilized Erchul and Marten's (2010)

integrated model of school-based consultation, whereby three interrelated tasks are simultaneously occurring. These tasks are the problem-solving task, the social influence task, and the support and development task. The tasks are considered interrelated because the problem-solving objectives of school-based consultation can only be accomplished through a social influence process between the consultant and the client, the goals of which are to assist the consultee improve his or her professional skills (Erchul & Martens, 2010).

The problem-solving task includes four stages: (1) problem identification, (2) problem analysis, (3) treatment implementation, and (4) treatment evaluation (Kratochwill & Bergan, 1990). In the problem identification stage, the teacher and consultant typically meet regarding specification of the problem. In the present study, however, the teacher did not initiate the consultation process, but instead volunteered to participate. Furthermore, the primary researcher and trained graduate students collected baseline data pertaining to the specified problem (i.e., disruptive student behavior and use of classroom management practices). The problem analysis stage was conducted in the same manner as traditional behavioral consultation (i.e., share data, discuss antecedents and consequences contributing to disruptive behavior, develop a plan utilizing the classroom management practices). The problem identification and problem analysis stages were conducted in order to establish rapport with each teacher during baseline meetings, provided a structure for meetings, and to inquire about teachers' current classroom management practices. Meetings were between 5-10 minutes in length. Additionally, observation times were determined during the problem identification stage

by asking teachers when they experience the most difficulty with classroom management (See Figure 6).

In the social influence task, methods of strategically influencing consultee behavior are utilized (Erchul & Martens, 2010). This was achieved in the present study through referent power and expert power. Referent power is influence based on a consultee's identification with a consultant/and or their desire for such identification (French & Raven, 1959; Raven, 1965, 1992, 1993). Expert power is influence based upon the perception the consultant possesses knowledge or expertise in the area. To promote referent power the consultant engaged in personal conversations at the start and end of consultation sessions and attempted to initiate in social interactions outside the classroom. Expert power occurred through presentation of information regarding the relationship between classroom management practices, student engagement, and disruptive behavior.

Several of Cialdini's (1993) social influence principles were also initiated during consultation. Building personal relationships with each teacher used the liking principle, whereby consultees prefer to acquiesce to the requests of people they know and like. The principle of consistency was also utilized. Once teachers committed to the study and subsequent meetings with the consultant, they experienced a sense of personal pressure to behave consistently with that commitment. Lastly, the contrasting principle (i.e., larger-then-smaller request) and reciprocation principle were used when suggesting changes to classroom practices. For instance, when asking teachers to create signs to hang throughout the classroom as cues for error correction and praise, the consultant first

suggested making notes in each day's lesson plan. Additionally, the consultant offered to make the signs for the teachers, which initiated a desire to repay in-kind the effort (i.e., reciprocation principle).

Plan implementation plus weekly performance feedback. In order to support and develop the teachers' ability to implement the plan and maintain treatment fidelity, performance feedback was added to the plan implementation stage (Erchul & Martens, 2010). Performance feedback was modeled after Noell et al. (2005) and consisted of the consultant meeting weekly with the teacher for 5 to 10 minutes and presenting direct observation classroom management and student engagement data. The data presented was on a graph created in Excel (i.e., percentage of intervals in which target behaviors were observed). The consultant reviewed data and discussed how classroom management and student behavioral data are connected (i.e., if the teacher increases use of classroom management practices, student engagement likely will increase). In addition, specific verbal praise was provided for treatment steps accurately completed as well as collaborative problem solving to improve implementation. Lastly, the performance feedback meeting ended with the consultant scheduling the next feedback meeting.

Problem evaluation. During the final performance feedback meeting, the teacher and consultant reviewed final data, examined plan effectiveness, and identified ways to ensure maintenance of effects. In addition, the social validity measure was administered to assess social validity from the teacher's perspective (Wolf, 1978). The teacher was informed at this meeting that the consultant would not be returning weekly. The teacher

was asked to continue using the intervention independently. The consultant had no further contact with the teacher.

Design

Single case design (SCD) is a well-established methodological approach for evaluating evidence-based practices in school psychology (Horner et al., 2005). SCD focuses on an individual "case" as the unit of intervention and unit of data analysis (Kratochwill et al., 2010). This case could be a single participant or a cluster of participants. The case provides its own control for purposes of comparison and allows for detailed analysis of responders or non-responders to an intervention. SCD is also feasible in practical settings and does not rely on random assignment as does group design (Horner et al., 2005).

To address the first three research questions a concurrent multiple baseline single case design was used. In a multiple baseline SCD, two or more baselines are concurrently established and the independent variable is sequentially introduced across the baselines (Kennedy, 2005). Multiple baseline designs are beneficial when the effects of the IV cannot be reversed once behavior is exposed to it. Additionally, due to the design logic of an A-B sequence for each baseline, such designs require fewer changes in educational settings than other $N = 1$ designs. The purpose of single case designs is to establish if there is a functional relationship between the independent and dependent variable (Horner et al., 2005). This occurs in multiple baseline designs when data collected during treatment demonstrate a pattern that differs significantly from data collected at baseline and is then replicated across subjects, behaviors, or settings (Horner et al., 2005).

Baseline data were collected for five data points or until a stable baseline trend was observed (Horner et al., 2005). Then treatment (performance feedback) was instituted at successive points in time for five teachers and data were collected three times a week for 10 weeks.

Criteria for designs that Meet Evidence Standards. According to Kratochwill and colleagues (2010) there are several criteria a SCD study must meet in order to *Meet Evidence Standards* according to What Works Clearinghouse. First, the independent variable must be systematically manipulated and each outcome variable measured systematically over time by more than one assessor. Inter-rater agreement needs to be measured in each phase or on at least 20% of the data points in each condition. The SCD study includes at least three attempts to demonstrate an intervention effect at three different time points or within three different phase repetitions. For a phase to be considered an attempt to demonstrate an intervention effect, the phase must have a minimum of three data points. In a multiple baseline there needs to be a minimum of six phases and at least five data points to *Meet Standards* (six phases with at least three data points per phase to *Meet Standards with Reservation*).

Analysis

In order to demonstrate a relationship between the IV and DV, there needs to be at least three indications of an intervention effect by documenting the consistency of level, trend, variability within each phase, as well as, the immediacy of effect, proportion of overlap, consistency of data points across phases, and comparing observed and projected patterns of the outcome variable (Horner et al., 2005; Kratochwill et al., 2010). Visual

analysis of results is considered the traditional approach to examining the functional relationship between and across phases (Horner et al., 2005). Kratochwill and colleagues (2010) suggest following four steps when examining within and between phase patterns. First, examine whether there is documentation of a predictable baseline pattern of data. Then evaluate data within each phase to assess within phase pattern(s), paying particular attention to whether there is sufficient data with sufficient consistency to demonstrate a predictable pattern (i.e., level, trend, and variability of data). Then compare data from each phase with adjacent phase to assess whether manipulation of the IV was associated with an effect (i.e., overlap, immediacy of effect, and consistency). Lastly, integrate all of the data from phases to determine whether there are at least three different demonstrations of an effect during at least three different time points (i.e., a functional relationship).

Within and between phase data patterns can be examined through six variables (Horner et al., 2005; Kratochwill et al., 2010). Within phase data patterns include level, which is the mean score for the data within a phase. While the absolute level within a phase is important, the last few data points contain the most essential information concerning the level of behavior before a phase change (Kennedy, 2005). Trend is the slope of the best fitting straight line for the data and variability is the range or standard deviation of data about the best fitting line. For between phase data patterns, the immediacy of effect or change in level between the last three data points in one phase and first three of the next phase can be examined. More rapid change provides more convincing evidence that the change in DV is due to the IV. Overlap is the proportion of

data from one phase that overlaps with data from a previous phase. The smaller the proportion of overlap, the more indicative of an effect. Finally, looking at data from all phases within the same condition and examining consistency in data patterns demonstrates the consistency of data in similar phases. The greater the consistency of data in similar phases, the more suggestive of a causal relationship.

Some researchers propose that visual analysis should be the primary method for single case design data (Brossart et al., 2006). They argue that visual analysis reveals any intervention effects large enough to be important and yields low error rates. In addition, they suggest visual analysis is conservative in identifying treatment effects and therefore the increased sensitivity that statistical analyses can offer is not necessary.

Despite these strengths, research has suggested the reliability of judgments made from visual analyses have consistently found to be low-to-moderate inter-rater reliabilities, in the range .40-.60 (Brossart et al., 2006). Trend lines have been suggested as a way to increase the reliability and validity of visual analysis, however, results have not been uniformly positive. Trend lines have been found to create dependencies, help maintain inconsistent judgments, and led to overestimates on trend to the neglect of other features. Another disadvantage of visual analysis is that subtle effects are difficult to detect, which often results in an increased probability of Type II error (Gresham & Vanderwood, 2008). Due to these disadvantages of visual analysis, including statistical analysis as a supplement is recommended (Brossart et al., 2006). Statistical procedures can be beneficial when there is not a stable baseline, expected treatment effects cannot be well predicted, and statistical control is necessary for extraneous factors. Therefore, in the

present study, effect size analysis was used in conjunction with visual analyses to address research questions 1-3.

Effect Size Analysis

Effect size analysis is a statistical method of examining the degree of the functional relationship between independent and dependent variables. The use of effect size analysis has not been customary in SCD research. Parker and colleagues (2007) found that of 75 multiple baseline designs reviewed, most researchers (87%) relied entirely on visual analysis. Due to movements for evidence-based interventions, practices, and treatments in education, however, more objective and reliable single case research results have been demanded (Parker et al., 2007). These changes have prompted the development of several methods for calculating SCD effect sizes. Within this context, a debate over the best effect size has ensued. The debate surrounds R^2, the regression effect size, and percent of non-overlapping data (PND). R^2 can be converted into Cohen's (1988) *d*, the standardized mean difference. PND is the calculation of non-overlap between baseline and successive intervention phases. PND can be hand-calculated by identifying the highest data point in baseline and determining the percentage of data points during intervention exceeding this level.

This debate highlighted several strengths and weaknesses associated with both R^2 and PND. As mentioned previously, R^2 can be converted to Cohen's *d* effect sizes, which are well established within the research community (Kratochwill et al., 2010; Parker et al., 2007). In addition, R^2 permits calculation of confidence intervals as an indicator of effect size reliability. Regression approaches also utilize all of the data in both phases in

SCD. Lastly, regression approaches can be more flexible in handling complex error structures and model phase trends. The debate determined at least three weaknesses of the regression approach. For instance, expertise is often needed to conduct and interpret regression analyses. Extreme outlier scores can also influence regression analyses and the parametric data assumptions of normality, equal variance, and serial independence are often not met by SCD data.

The competing approach, PND, demonstrates three advantages. First, PND can be easily calculated, with a pencil and a ruler on a printed graph, and as a percentage calculation (Kratochwill et al., 2010; Parker et al., 2007). Similarly, PND is acceptable to visual analysts, as PND's emphasis on overlapping data reflects a key feature of most visual analyses. The third advantage is PND's applicability to any SCD. As with regression approaches, however, PND has limitations. PND is not considered an effect size nor related to an accepted effect size, therefore it needs its own interpretation guidelines. In addition, PND has unknown reliability and lacks a known sampling distribution, so *p* values and confidence intervals cannot be calculated. Ignoring all baseline data except one data point, which due to its extremity is likely to be unreliable, is another weakness of PND. Finally, PND lacks sensitivity or discrimination ability as it nears 100% for very successful interventions.

Due to these limitations, researchers have proposed another index, the "percent of all non-overlapping data" (PAND; Parker et al., 2007). PAND is calculated by dividing the total number of intervention data points that overlapped baseline by the number of all data points (intervention and baseline combined). Similar to PND, PAND reflects data

non-overlap between phases, however PAND uses data from both phases. This helps eliminate the criticism of overemphasis on one unreliable data point. PAND is closely related, via a 2 X 2 table, to the Pearson's *Phi* effect size (Parker et al., 2007; Riley-Tillman & Burns, 2009). This effect size represents the difference between the average level of the intervention phase and the average level of a baseline phase divided by the pool standard deviation. On its own PAND lacks status, however, *Phi* has known sampling distributions, so *p* values are available, statistical power can be estimated, and confidence intervals can be included to indicate effect size reliability.

The data requirements for PAND are also minimal. A minimum of 20 data points are necessary and it is not subject to the parametric assumptions of normality and equal variance (Parker et al., 2007). Studies have also indicated that sufficient statistical power is easy to attain. Parker and colleagues (2007) found most multiple baseline designs that contained 45 to 96 data points (relatively balanced between phases) and effect sizes of moderate magnitude had sufficient power. Using data from Parker's study, one could calculate PAND. For instance, if visual analysis yielded overlapping data as 2 for "Adam," 2 for "Bob," and 2 for "Carol," totaling 6, or 6/28 = 21.4% overlap. PAND is subsequently, 100 – 21.4% = 78.6%. Phi is then calculated by creating a 2 x 2 table with the data:

Cell A: % of baseline (baseline/total)	Cell B: % overlapping data/2
Cell C: % overlapping data/2	Cell D: % of tx (tx/total)

After constructing the table, the following equation is calculated: $\Phi = [a / (a + c)] - [b / (b + d)]$ (Parker, Hagan-Burke, & Vannest, 2007). Continuing from the previous example, the percentage of data points in the baseline and intervention phases are calculated: 13/28 = 46.4%, 15/28 = 53.6%. These percentages are then entered into their respective columns. Next, the proportion of overlapping data (21.4%) is split between cells B and C: 10.7% in each cell. These cells indicate "too high" scores in the baseline phase (cell B) and "too low" scores in the intervention phase (cell C). Lastly, cells A and D are calculated by subtraction: 46.4 - 10.7 = 35.7 and 53.6 – 10.7 = 42.9. Applying the above formula, $\Phi = 35.7/46.5 – 10.7/53.6 = .768 - .199 = .569$ or .57. Using Cohen's (1988) guidelines, phi coefficients of .10 are considered small, .30 are medium, and .50 are large. Therefore, the above Phi of .57 would indicate a large effect.

Cell A: 35.7%	Cell B: 10.7%
Cell C: 10.7%	Cell D: 42.9%
Total: 46.4%	Total: 53.6%

While using PAND/Phi in single case designs can remedy some of the deficiencies of PND and R^2, it is not without its limitations. For instance PAND lacks sensitivity at the upper end of the scale. Similar to PND, when there is no data overlap between Phases A and B, PAND awards a 100% score, regardless of the distance between the data clusters. In addition, PAND also measures simple mean level shifts, not accounting for positive baseline trend. In order to infer a causal link between intervention and behavior, a positive baseline trend must be considered. Large effect sizes alone do

not indicate that change was due to the intervention. Therefore, a combination of visual analysis and effect sizes was utilized in analysis. Hence, one can visually assess and confirm if effect sizes are an accurate depiction of the data.

Threats to Validity

Multiple baseline designs have certain advantages, such as being more ethically desirable due to intervention not being withdrawn (Kennedy, 2005). However, this can also be considered a weakness in that a return to baseline is seen as necessary to prove cause-and-effect. To remedy this issue, many replications of the study are needed to substantiate the suggested relationship. Another concern is the prolonged nature of baselines when intervention is needed immediately and therefore an alternating treatment design may be a better approach. In addition to the concerns related to multiple baseline designs, possible threats to validity can occur in any single case design. Internal validity pertains to the validity of inferences about whether observed covariation between treatment and outcome reflects a causal relationship (Shaddish et al., 2002). Possible threats to the internal validity of the present study included selection bias, history, maturation, attrition, instrumentation, and interaction among attrition and other threats. Selection bias is differences in respondent characteristics. History pertains to events occurring concurrently with treatment and maturation is naturally occurring changes over time. Attrition is the potential loss of participants. Instrumentation refers to the nature of a measure. Lastly, interaction among attrition and other threats is where one of the previously mentioned threats influences participants not to complete the study (Kennedy, 2005). Internal validity can be improved through replication and randomizations, as well

as at least three demonstrations of an effect at three different time points (Horner et al., 2005). External validity is the extent to which findings can be generalized to other participants, places, or conditions (Horner et al., 2005). This is achieved through direct and systematic replication. However, external validity can be restricted by selection bias and attrition, limiting the examples available for analysis. Alternatively, external validity can be enhanced through replications across different subjects, conditions, and dependent variables. External validity can also be improved by providing operational definitions of participants, places, and conditions that influenced behavior prior to treatment (i.e., thorough description of baseline). Horner and colleagues (2005) also recommend evaluating social validity when using single case designs. A typical intervention agent, in a typical context, as described previously, can enhance social validity, through implementation of the independent variable over an extended time period.

Each of these possible threats to validity was addressed. Although participating teachers represented a convenience sample, a thorough description of participants was provided in order to determine their relative equivalence. To address history effects, possible extraneous variables were documented and noted as possible limitations. Since the study occurred within a short time frame, maturation effects are unlikely. Seeking five teachers to participate, as well as adding an incentive to increase the likelihood of continued participation, weakens the possibility of attrition. By ensuring observers were properly trained, having adequate inter-rater agreement, and providing clear operational definitions, instrumentation was not a threat. Since the possibilities of other internal validity threats were addressed, an interaction among attrition and other threats was also

reduced. To improve upon the study's external validity, thorough operational definitions of the participants, conditions, and dependent variables were provided. Furthermore, social validity data were collected and consultation plus performance feedback was provided by a typical intervention agent in a typical setting, enhancing social validity.

Results

Research Question 1

Research question 1 asked: To what extent is there a functional relationship between performance feedback with a teacher and student engagement? Figure 7a suggests that all teachers had a positive effect on level of total engagement and changes in trend. Immediacy was seen with Teachers 1, 4, and 3. Teacher 5's demonstrated a delayed increase (after 2 treatment sessions), at which point the level increased. Baseline and intervention phases for Teachers 1, 5, 4, and 3 had no overlapping data points. A decrease in variability was seen for each teacher and data patterns demonstrated relative consistency in intervention phases. Effect size analysis for total engagement indicated a large effect. PAND for total student engagement was 98.5%. This resulted in a Phi of .97.

Figures 7b and 7c shows no observable changes in level, trend, immediacy, and variability was seen for active and passive engagement when graphed separately. Teacher 5 showed a slight decrease in variability for both after first application of performance feedback but no change in trend or level. PAND for active engagement was 65.2%, resulting in a Phi of .30, indicating a medium effect size. PAND for passive engagement was 65.9%, resulting in a Phi of .31, indicating a medium effect size.

Research Question 2

Research question 2 asked: To what extent is there a functional relationship between performance feedback with a teacher and disruptive student behavior? Figure 7d suggests that all teachers had a negative effect (i.e., decrease) on level of disruptive behavior. Teacher 5's data did indicate a slight negative trend at baseline, however, there was also a decrease in variability for the intervention phase. Immediacy was seen with Teachers 1 and 3, as well as negative slopes of low-magnitudes in the intervention phase. The data patterns for Teachers 4 and 3 demonstrated relative consistency in similar phases. For disruptive student behavior PAND was 83.7% resulting in a Phi of .67, indicating a large effect size.

Research Question 3

Research question 3 asked: To what extent is there a functional relationship between performance feedback with a teacher and teachers' implementation fidelity of targeted classroom management practices. Targeted classroom management practices included specific praise, error correction, prompts/pre-corrections, active supervision, and providing OTRs. Each practice was examined first using visual analysis and then with effect size analysis.

Specific Praise. Figure 7e suggests that all teachers had a positive effect on level of specific praise. Immediacy was seen with Teachers 1, 5, 2, and 3, as well as a change in trend. Positive trends of low to medium-magnitudes were demonstrated for each. Teacher 4's demonstrated a delayed increase (after 2 treatment sessions), at which point the level increased markedly. Teacher 4 also indicated a positive trend with a high-

magnitude. Baseline and intervention phases for Teachers 1, 5, and 3 yielded no overlapping data points. The data patterns for Teachers 5, 4, and 3 demonstrated relative consistency of data in similar phases. Effect size analysis indicated a large effect. PAND for specific praise was 95.3%. This resulted in a Phi of .91.

Error Correction. Figure 7f shows that Teachers 1, 5, and 4 had a positive effect on level of error correction. Immediacy was seen with Teachers 1 and 5. Teacher 4's demonstrated a delayed increase (after 2 treatment sessions), at which point the level increased markedly. Change in trend was demonstrated for Teachers 1, 5, and 4. A positive trend of medium-magnitude was indicated for Teacher 5 and a positive trend of low-magnitude for Teachers 1 and 4. Effect size analysis indicated a large effect. PAND for error correction was 84.5%. This resulted in a Phi of .69.

Prompts/Pre-Corrections. Figure 7g shows that there was a slight positive effect on level of prompts/pre-corrections for Teachers 1 and 5. Immediacy was seen with Teacher 1 in regards to level increase, as well as an initial medium positive slope, then a curvilinear data pattern (i.e., U pattern). No observable changes in level, trend, immediacy, and variability was seen for Teachers 2 and 3. Teacher 4 showed a decrease in variability after first application of performance feedback but no change in trend or level. Effect size analysis of prompts/pre-corrections indicated a small effect. Percent of all non-overlapping data (PAND) points was 55.0%. This resulted in a Phi of .10 (See Table 4 for all effect size statistics).

Active Supervision. Figure 7h indicates that Teacher 5 had a slight positive effect on level of active supervision. Immediacy was also seen with this teacher in regards to

level increase, as well as a change in trend. A positive trend with a low-magnitude slope was demonstrated in the intervention phase. No observable changes in level, trend, immediacy, and variability was seen for the remaining four teachers. Effect size analysis indicated a medium effect. PAND for active supervision was 65.9%. This resulted in a Phi of .31.

Opportunities to Respond. Figure 7i shows no observable changes in level, trend, immediacy, and variability was seen for the five teachers. Effect size analysis indicated a small effect. PAND for OTRs was 52.7%, resulting in a Phi of .05.

Research Question 4

Research question 4 asked: To what extent do teachers find performance feedback as addressing socially significant goals, socially acceptable procedures, and socially important outcomes? All teachers completed a 15-item social validity measure (See Figure 2), with each question rated on a 1(strongly disagree) to 6 (strongly agree) scale. Overall, teachers' social validity values ranged from 71 (SD = 0.7) to 90 (SD = 0), with a mean of 84 (See Table 5 for each teacher's total) indicating teachers found the intervention as socially valid. They also provided ratings that the performance feedback procedures used would be acceptable for their school (M = 5.6, SD = .89), most teachers would find performance feedback appropriate (M = 5.4, SD = .89), suitable for the described/stated purposes (increasing student engagement; M = 5.4, SD = .89), and performance feedback should prove effective in meeting the purposes (M = 5.6, SD = .89). Teachers also would suggest the use of performance feedback to other teachers (M = 5.6, SD = .89), and that it is a fair (M = 5.6, SD = .55) and appropriate to meet the

school's needs (M = 5.8, SD = .45). Teachers also indicated that performance feedback would not result in negative side effects for the students (M = 6; SD = 0), would be appropriate for a variety of students (M = 6, SD = 0), and is consistent with those I have used in school settings (M = 5, SD = .55). Lastly, teachers liked the procedures used in performance feedback (M = 5.6, SD = .89), would be willing to use it in the school setting (M = 5.4, SD = .89), and overall considered performance feedback as beneficial for elementary school students (M = 5.8, SD = .45).

Discussion

The purpose of the present study was to examine the relationship between performance feedback and teacher and student behavioral outcomes. This study advanced the performance feedback literature by addressing still unanswered questions. The aforementioned was achieved by utilizing training that aligned with traditional consultation, following a standardized performance feedback protocol (Fallen et al., 2015; Noell et al., 2005), and providing weekly performance feedback using graphs of teacher and student data. Furthermore, the current study addressed previous limitations by using methodology that meets WWC *Meets Evidence Standards* and collected adequate treatment and inter-rater agreement data. This study also demonstrated the effectiveness of performance feedback within a school setting. Most importantly, this research indicated whether performance feedback with a teacher could decrease disruptive student behavior and increase student academic engagement as well as increase teachers' treatment integrity of two evidence-based classroom management practices. Lastly, this study determined teachers found performance feedback procedures

as socially valid.

Performance Feedback and Student Engagement

Research question 1 explored whether there was a functional relationship between performance feedback and student engagement. Results indicated total engagement increased across teachers. A positive effect on level was seen for all teachers, as well as a change in trend. Immediacy was seen with three teachers and no overlapping data points for four teachers. Each teacher had a positive trend of low to medium-magnitude. A decrease in variability was seen for each teacher and data patterns demonstrated relative consistency in intervention phases. Effect size analysis for total engagement indicated a large effect ($\Phi = .97$). Lastly, all participant's total engagement went from below 80% in baseline to 80% and above in the intervention phase. Research suggests that effectively managed classrooms have a rate of 80% academic engaged time or higher (Rathvon, 2008).

The results were less clear regarding specific areas of engagement. There were no observable changes in level, trend, immediacy, and variability for active and passive engagement when graphed separately. One teacher showed a slight decrease in variability for both after first application of performance feedback but no change in trend or level. A medium effect size was generated for both ($\Phi = .30$; $\Phi = .31$). These findings may have been impacted by the nature of the lesson being observed, whereby some produced more active student behaviors and others more passive behaviors (Shapiro, 2004). Hence, total engagement may be more reflective of the relationship between performance feedback and student engagement.

Performance Feedback and Disruptive Student Behavior

The second research question in this study examined if there was functional relationship between performance feedback and disruptive student behavior. Results suggest that all teachers had a negative effect (i.e., decrease) on level of disruptive behavior. Immediacy was seen with two teachers, as well as changes in trend. The data patterns for two teachers demonstrated relative consistency in similar phases. Effect size analysis yielded a large effect size ($\Phi = .67$).

The decreases in disruptive behavior and increases in student engagement may also correspond to the improved fidelity of specific praise and error correction. As mentioned previously, specific praise and error correction are two methods that can be used to increase engagement and reduce disruptive behavior (Simonsen et al., 2008). These practices also demonstrated significant effects in the present study. Engagement has also been found to be the best mediating variable between instruction and academic achievement whereby if students are actively engaged in instruction, then it is difficult to engage in incompatible behavior (e.g., talking out; Greenwood et al., 2002). However, due to the type of analysis conducted, only a hypothesis of potential connection can be made rather than any causality.

Performance Feedback and Classroom Management

Research question 3 examined the impact of performance feedback on teachers' implementation fidelity of targeted classroom management practices. Results indicated that specific praise and error correction were the only practices enhanced by performance feedback. Visual analysis results demonstrated no relationship between performance

feedback and prompts/pre-corrections, active supervision, and opportunities to respond. Effect size analysis suggested small to medium effects for these practices, however, due to limitations in effect size analysis within single case design, especially a lack of control for baseline trend, only the effect sizes for areas meeting visual analysis standards were considered further.

All five teachers increased their use of specific praise in the classroom. Immediacy and a change in trend were indicated with four teachers. Baseline and intervention phases for three teachers yielded no overlapping data points. The data patterns also demonstrated relative consistency of data in similar phases. Furthermore, effect size analysis yielded a large effect ($\Phi = .91$). In regards to error correction, three teachers increased their use. Immediacy was seen with two teachers and a delayed increase with another. Change in trend was seen for all three teachers. Effect size analysis indicated a large effect ($\Phi = .69$).

Social Validity

Lastly, research question 4 pertained to whether teachers found performance feedback as addressing socially significant goals, socially acceptable procedures, and socially important outcomes. All teachers completed a 15-item social validity measure (See Figure 2), with each question rated on a 1(strongly disagree) to 6 (strongly agree) scale. Teachers found performance feedback as suitable for the stated purposes, liked the procedures used, and considered it as beneficial for elementary school students. In general, results suggested teachers found the intervention as socially valid.

Overall, this study demonstrated a possible functional relationship between

performance feedback with two of the classroom management practices and total student engagement and disruptive behavior. Possible explanations for the increase in the practices (i.e., specific praise and error correction) are the relative simplicity of implementation, and the previously established relationship between these practices and student engagement and disruptive behavior (Simonsen et al., 2008). While these findings are important to note, the non-effects are also worth analyzing. The exploration of non-effects may indicate other potential moderators impacting the effectiveness of performance feedback in the classroom.

Discussion of Non-Effects

According to Gresham (1989), treatment integrity can be influenced by intervention complexity, number of facets or components, and time required. As discussed previously, specific praise and error correction are among the simplest and most empirically validated classroom management strategies to implement (Simonsen et al., 2008). Due to the 10-week timeframe, teachers may have been easily able to incorporate specific praise and error correction into their classroom routine, whereas the other practices could have been more complex. Subsequently, prompts/pre-corrections, active supervision, and OTRs may have presented as too time consuming or challenging to learn and implement quickly. These practices may have required more in-depth training, including modeling and practice when students were present in order to produce effects.

Furthermore, the present study represents one of the first to include five classroom management practices as variables. Previous studies only attempted to

improve the treatment fidelity of three or fewer practices (Reinke et al., 2008; Simonsen et al., 2010). For instance, Rienke and colleagues (2008) only utilized performance feedback to increase the treatment integrity of specific praise. Similarly, Simonsen et al. (2010) addressed teachers' implementation fidelity of three practices (i.e., prompts/pre-corrections, OTRs, and specific praise). Results were similar to the present study, whereby the most notable and sustained behavior change occurred for specific praise. Teachers' integrity of prompts/pre-corrections and OTRs remained relatively low. Anecdotally, teachers also reported specific praise resulted in the "most improvement in [their] students," "feels the most normal," and that prompts/pre-corrections and OTRs "remained as difficult skills." Therefore, attempting to enhance teachers' treatment integrity of five classroom management practices may not be advisable and prompts/pre-corrections, active supervision, and OTRs possibly represent complex practices and consequently require more time for development. Additionally, teacher behaviors were recorded using partial interval recording, which may have contributed to low levels of observed behavior. Hence, a frequency count may have provided a more accurate measure of teacher behavior.

Future Research

Future research should examine how performance feedback can be effectively utilized to enhance the treatment integrity of more than three classroom management practices. For instance, the use of other single-case designs (e.g., alternating treatment) may better address the relationship between performance feedback and five classroom management practices. Through an alternating treatment design, each classroom

management practice could be evaluated by rapidly interchanging interventions in a random sequence (Kennedy, 2005). The timeframe (i.e., 10-weeks) could also be expanded allowing the consultant to stagger presentation of each classroom management practice and the teachers more time to develop each skill.

Noell and colleagues (1997) findings indicated a less rigorous training phase could be used prior to implementation and treatment integrity still improved. Similarly, Mortensen and Witt (1998) suggest that, while effects may not be as high, weekly rather than daily performance feedback meetings can be effective. However, if prompts/pre-corrections, active supervision, and OTRs represent more difficult practices to develop, rigorous training and daily performance feedback may be necessary for improved outcomes. Additionally, previous studies examining performance feedback and classroom management practices (Reinke et al., 2008; Simonsen et al., 2010) provided thorough trainings (e.g., explicit training, discussion, activities, and self-management strategies) prior to implementation, as well as daily feedback. Therefore, future research exploring the amount of training and performance feedback necessary in order for teachers to enhance their treatment integrity of these practices may be warranted. Lastly, as advances in technology continue, the notion of incorporating videos of teachers' use of the practices into performance feedback sessions represents an interesting direction for the performance feedback literature. The consultant could videotape the teachers' use of practices during observation sessions and subsequently provide feedback based on these videos. The impact of performance feedback utilizing videos versus without could then be compared.

Limitations

As with every study, there are several limitations worth noting. A major limitation is the lack of a steady baseline in some areas. Due to practical time constraints, it was not possible to wait until a stable baseline was present. However, at least five data points were present within each phase (Horner et al., 2005). Another limitation is that no maintenance phase was conducted. Without a maintenance phase, it cannot be determined whether effects would be sustained after discontinuance of performance feedback. Additionally, reactivity to being monitored and seeing data on treatment integrity may have been adequate to initiate change in the teachers' behavior independent of the performance feedback procedures.

Another limitation is reliability of the observation instrument. Due to the instrument being created by the primary researcher, it has not undergone rigorous study to establish reliability and validity. This may produce a threat to the study's internal validity whereby the nature of the measure (i.e., instrumentation) impacts the validity of the causal relationship (Shaddish et al., 2002). Fortunately, observers were properly trained, clear operational definitions included, and pilot testing was conducted to ensure the observational coding scheme described applies to participants in this study (Horner et al., 2005). The inter-rater reliabilities for raters ranged from .94-1.00 as calculated by Cohen's kappa. These reliabilities are well above the minimum acceptable value of .60 suggested by Hartmann, Barrios, and Wood (2004). Content validity was also established by having advanced graduate students not involved in development determine whether the measure adequately sampled the domain of interest (Raykov & Marcoulides, 2011).

Additionally, inter-rater reliabilities for raters ranged from .95-1.00 during the present study. Future studies on the reliability of this instrument in different settings are warranted, however, to further establish if it has adequate reliability.

Implications

Teachers are increasingly faced with challenging student behavior, which limits time devoted to instruction and can lead to negative student outcomes. Unfortunately, traditional professional development models are often ineffective at preparing teachers to handle disruptive student behavior (Fixsen, Naoom, Blasé, Freidman, & Wallace, 2005; Kinkead, 2007). As research suggests, one-time consultation and training is not enough to effectively create classroom change. Alternatively, performance feedback has been found to increase teachers' use of evidence-based classroom management practices, which can subsequently reduce disruptive student behavior and increase student engagement (Reinke et al., 2008). This is especially important given that research suggests classrooms with poor behavior management place students at-risk for future behavior problems (Kellam et al., 1998).

This study generated similar results, whereby performance feedback enhanced teachers' treatment integrity of specific praise and error correction and increased student engagement and decreased disruptive behavior. Such findings suggest performance feedback may be an efficient method for improving classroom management at a broader scale. Instead of resources being allocated to one individual student at a time, the academic engagement of an entire classroom of students can be improved at once. Furthermore, if students are engaged, then it is difficult to engage in disruptive behavior,

providing more time for instruction and subsequently improving academic achievement (Greenwood et al., 2002). Results also expanded the literature on school-based consultation. The study used training that aligned with traditional consultation, followed performance feedback as previously described, and provided weekly performance feedback. Additionally, methodology met WWC *Meets Evidence Standards* and yielded adequate treatment and inter-rater agreement data. The methodology of the previous studies evaluating performance feedback and classroom management practices would not meet WWC *Meets Evidence Standards* and often lacked treatment integrity, adequate inter-rater agreement, and teacher and student outcome data.

References

Allen, J., Pianta, R., Gregory, A., Mikami, A., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science, 333*, 1034-1037.

Becker, K. D., Bradshaw, C. P., Domitrovich, C., & Ialongo, N. S. (2013). Coaching teachers to improve implementation of the good behavior game. *Administration and Policy in Mental Health, 40*(6), 482-493.

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

Bradshaw, C. P., Pas, E. T., Goldweber, A., Rosenberg, M., & Leaf, P. (2012). Integrating school-wide positive behavioral interventions and supports with tier 2 coaching to student support teams: The PBISplus model. *Advances in School Mental Health Promotion: Training and Practice, Research and Policy, 5*, 177-193.

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.

Brown, J. L., Jones, S. M., LaRusso, M. D., & Aber, J. L. (2010). Improving classroom quality: Teacher influences and experimental impacts of the 4R program. *Journal of Educational Psychology, 102*, 153-167.

Burke, R. V., Oats, R. G., Ringle, J. L., Fichtner, L., & DelGaudio, M. B. (2011). Implementation of a classroom management program with urban elementary schools in low-income neighborhoods: Does program fidelity affect student behavior and academic outcomes? *Journal of Education for Students Placed at Risk, 16*, 201-218.

California Department of Education Data Reporting Office. (2014). Retrieved on July 30th, 2014 from http://data1.cde.ca.gov/dataquest.

Cameron, J., & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research, 64*(3), 363-423.

Cappella, E., Hamre, B., Kim, H., Henry, D., Frazier, S., Atkins, M., & Schoenwald, S. (2012). Teacher consultation and coaching within mental health practice: Classroom and child effects in urban elementary schools. *Journal of Consulting and Clinical Psychology, 80*, 597-610.

Cappella, E., Jackson, D., Wagner, C., Hamre, B. K., Soulé, C. (2011). Bridging mental health and education in urban elementary schools: Participatory research to inform intervention development. *School Psychology Review, 40*(4), 486-508.

Carr, E. G., Taylor, J. C., & Robinson, S. (1991). The effects of severe behavior problems in children on the teaching behavior of adults. *Journal of Applied Behavioral Analysis, 24*(3), 523-535.

Cialdini, R. (1993). *Influence: Science and practice* (3rd ed.). New York: Harper Collins.

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

Colvin, G., Sugai, G., Good, R. H., III, & Lee, Y. (1997). Using active supervision and precorrection to improve transition behaviors in an elementary school. *School Psychology Quarterly, 12,* 344-363.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology, 18,* 105-115.

Deci, E. L. (1976). Notes on the theory and meta-theory of intrinsic motivation. *Organizational Behavior and Human Performance, 15*(1), 130-145.

Denton, C. A., & Hasbrouck, J. (2009). A description of instructional coaching and its relationship to consultation. *Journal of Educational & Psychological Consultation, 19*, 150-175.

De Pry, R. L., & Sugai, G. (2002). The effect of active supervision and pre-correction on minor behavioral incidents in a sixth grade general education classroom. *Journal of Behavioral Education*, *11*, 255-264.

Dishion, T., Spracklen, K., Andrews, D., & Patterson, G. (1996). Deviancy training in male adolescent friendships. *Behavior Therapy, 27,* 370-390.

Driscoll, K. C., & Pianta, R. C. (2010). Banking time in Head Start: Early efficacy of an intervention designed to promote supportive teacher-child relationships. *Early Education and Development, 21*, 38-64.

DuPaul, G. J., Volpe, R. J., Jitendra, A. K., Lutz, J. G., Lorah, K. S., & Gruber, R. (2004). Elementary school students with AD/HD: Predictors of academic achievement. *Journal of School Psychology, 42*, 285-301.

Eddy, J. M., Leve, L., & Fagot, B. (2001). Coercive family processes: A replication and

extension of Patterson's coercion model. *Aggressive Behavior, 27*, 14-25.

Elam, S. M., Rose, L. C., & Gallup, A. M. (1996). The 28th annual Phi Delta Kappa/Gallup poll of the public's attitudes toward the public schools. *Phi Delta Kappan,* 41-58.

Erchul, W. P., & Martens, B. K. (2010). *School consultation: Conceptual and empirical bases of practice- 3rd edition*. New York, NY: Springer.

Erchul, W., & Sheridan, S. (2008). *Handbook of Research in School Consultation*. New York, NY: Lawrence Erlbaum Associates.

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

Fixsen, D., Naoom, S., Blase, K., Friedman, R., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI Publication No. 231). Tampa: University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network.

Forman, S. G., Olin, S., Hoagwood, K., Crowe, M., & Saka, N. (2009). Evidence-based intervention in schools: Developers' views of implementation barriers and facilitators. *School Mental Health, 1*, 26-36.

French, J. R. P., Jr., & Raven, B. H. (1959). The bases of social power. In D. Cartwright (Ed.), *Studies in Social Power* (pp. 150-167). Ann Arbor, MI: Institute for Social Research.

Frank, J. L., & Kratochwill, T. R. (2008). School based problem-solving consultation: Plotting a new course for evidence-based research and practice in consultation. In W. Erchul, & S. Sheridan (Eds.), *Handbook of Research in School Consultation* (pp. 13-30). New York, NY: Lawrence Erlbaum Associates.

Fuchs, L. S., & Fuchs, D. (1989). Exploring effective and efficient prereferral interventions: A component analysis of behavioral consultation. *School Psychology Review, 18*, 260-283.

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

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

Gresham, F. M., Gansle, K., & Noell, G. H. (1993). Treatment integrity in applied behavior with children. *Journal of Applied Behavior Analysis, 26*, 257-263.

Gresham, F. L., & Vanderwood, M. (2008). Quantitative research methods and designs in consultation. W. Erchul, & S. Sheridan (Eds.), *Handbook of Research in School Consultation* (pp. 63-88). New York, NY: Lawrence Erlbaum Associates.

Hartmann, D. P., Barrios, B. A., & Wood, D. D. (2004). Principles of behavioral observation. In S. N. Haynes, & E. M. Hieby (Eds.), *Comprehensive handbook of psychological assessment (Vol. 3, Behavioral assessment*) (pp. 108-127). New York, NY: John Wiley & Sons.

Hershfeldt, P. A., Pell, K., Sechrest, R., Pas, E. T., & Bradshaw, C. P. (2012). Lessons learned coaching teachers in behavior management: The PBISplus coaching

model. *Journal of Educational and Psychological Consultation, 22*, 280-299.

Hintze, J. M. (2005). Psychometrics of direct observation. *School Psychology Review, 34*, 507-519.

Hintze, J. M., Volpe, R. J., & Shapiro, E. S. (2008). Best practices in the systematic direct observation of student behavior. In A. Thomas, & J. Grimes (Eds.), *Best Practices in School Psychology* (pp. 319-336). Bethesda: The National Association of School Psychologists.

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*(2), 165-179.

Johnson, J., Duffett, A., Vine, J., & Moye, L. (2003). *Where are we now: 12 things you need to know about public opinion and public schools*. Retrieved December 17, 2013, from http://www.publicagenda.org/.

Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology, 10,* 165-185.

Kennedy, C. H. (2005). *Single-case designs for educational research*. New York, NY: Allyn and Bacon.

Kinkead, S. (2007). *Improving instruction through coaching.* Olympia, WA: The Center For Strengthening the Teacher Profession, Superintendent of Public Instruction.

Kratochwill, T. R., & Bergan, J. R. (1990). *Behavioral consultation in applied settings:*

An individual guide. New York, NY: Plenum.

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

Le Blanc, M., Vallieres, E., & McDuff, P. (1992). Adolescent's school experience and self-reported offending: An empirical elaboration of an interactional and developmental school social control theory. *International Journal of Adolescence and Youth, 3*, 197-247.

Lepper, M. R., & Greene, D. (1975). Turning play into work: Effects of adult surveillance and extrinsic rewards on children's intrinsic motivation. *Journal of Personality and Social Psychology, 31*, 479-486.

Lewis, R. (2001). Classroom discipline and student responsibility: The students' view. *Teaching and Teacher Education, 17*, 307-319.

Lewis, T. J., Colvin, G., & Sugai, G. (2000). The effects of pre-correction and active supervision on the recess behavior of elementary students. *Education and Treatment of Children, 23*(2), 109-121.

MacLeod, I. R., Jones, K. M., Somers, C. L., & Havey, J. M. (2001). An evaluation of the effectiveness of school-based behavioral consultation. *Journal of Educational and Psychological Consultation, 12*, 203-216.

MacSuga, A. S., & Simonsen, B. (2011). Increasing teachers' use of evidence-based classroom management strategies through consultation: Overview and case studies. *Beyond Behavior, 20*, 4-12.

Maggin, D. M., Swaminathan, H., Rogers, H. J., O'Keeffe, B. V., Sugai, G., & Horner, R. H. (2011). A generalized least squares regression approach for computing effect sizes in single-case research: Application examples. *Journal of School Psychology, 49*, 301-321.

Markow, D., Moessner, C., & Horowitz, H. (2006). *The MetLife survey of the American teacher*. Retrieved December 14, 2013, from http://www.metlife.com/WPSAssets/818214027011605058 71V1F2006MetLifeTeacherSurvey.pdf.

Matheson, A. S., & Shriver, M. D. (2005). Training teachers to give effective commands: Effects on student compliance and academic behaviors. *School Psychology Review, 34*(2), 202-219.

McEvoy, A., & Welker, R. (2000). Antisocial behavior, academic failure, and school climate: A critical review. *Journal of Emotional and Behavioral Disorders, 8*, 130-140.

McIntyre, E., Kyle, D., Chen, C., Munoz, M., & Beldon, S. (2010). Teacher learning and ell reading achievement in sheltered instruction classrooms: Linking professional development to student development. *Literacy Research and Instruction, 49*(4), 334-351.

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.

Mortenson, B. P., & Witt, J. C. (1998). The use of weekly performance feedback to increase teacher implementation of a prereferral academic intervention. *School Psychology Review, 27*(4), 613-627.

Myers, D. M., Simonsen, B., & Sugai, G. (2011). Increasing teachers' use of praise with a response-to-intervention approach. *Education and Treatment of Children, 34*(1), 35-59.

National Association of State Boards of Education. (1997). Schools without fear (Report of the NASBE study group on violence and its impact on schools and learning). Alexandria, VA: Author.

Nelson, J. R. (1996). Designing schools to meet the needs of students who exhibit disruptive behaviors. *Journal of Emotional and Behavioral Disorders, 4*, 147-161.

Nelson, J. R., Martella, R. M., & Marchand-Martella, N. (2002). Maximizing student learning: The effects of a comprehensive school-based program for preventing problem behaviors. *Journal of Emotional and Behavioral Disorders*, *10*, 136-148.

Noell, G. H. (2008). Research examining the relationships among consultation process, treatment integrity, and outcomes. In W. Erchul, & S. Sheridan (Eds.), *Handbook of Research in School Consultation* (pp. 323-342). New York, NY: Lawrence Erlbaum Associates.

Noell, G. H., & Gresham, F. M. (1993). Functional outcome analysis: Do the benefits of

consultation and prereferral intervention justify the costs? *School Psychology Quarterly*, *8,* 200-226.

Noell, G. H., & Witt, J. C. (1996). A critical re-evaluation of five fundamental assumptions underlying behavioral consultation. *School Psychology Quarterly, 11*(3), 189-203.

Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L., Koenig, J. L.,…Duhon, G. J. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review, 34*, 87-106.

Noell, G. H., Duhon, G. J., Gatti, S. L., & Connell, J. E. (2002). Consultation, follow-up, and implementation of behavior management interventions in general education. *School Psychology Review, 31*(2), 217-234.

Noell, G. H., Witt, J. C., LaFleur, L. H., Mortenson, B. P., Rainer, D. D., & LaVelle, J. (2000). Increasing intervention comparison of two follow-up strategies. *Journal of Applied Behavioral Analysis, 33*, 271-284.

Noell, G. H., Witt, J. C., Gilbertson, D. N., Rainer, D. D., & Freeland, J. T. (1997). Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *School Psychology Quarterly, 12*(1), 77-88.

Noll, M. B., Kamps, D., & Seaborn, C. F. (1993). Prereferral intervention for students with emotional or behavioral risks: Use of a behavioral consultation model. *Journal of Emotional and Behavioral Disorders, 1,* 203–214.

Ota, K. R., & DuPaul, G. J. (2002). Task engagement and mathematics performance in children with attention deficit hyperactivity disorder: Effects of supplemental computer instruction. *School Psychology Quarterly, 17*, 242-257.

Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping data (PAND): An alternative to PND. *Journal of Special Education, 40*(4), 194-204.

Patterson, G. R., Reid, J., & Dishion, T. (1992). *Antisocial boys*. Eugene, OR: Castalia.

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

Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K. (2008a). Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 23*, 431-451.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008b). *Classroom Assessment Scoring System [CLASS] Manual: Pre-K*. Baltimore, MD: Brookes Publishing.

Rathvon, N. (2008). *Effective school interventions: Evidence-based strategies for improving student outcomes.* New York, NY: Guilford.

Raven, B. H. (1965). Social influence and power. In I. D. Steiner, & M. Fishbein (Eds.), *Current Studies in Social Psychology*. New York: Holt, Rinehart, Winston.

Raven, B. H. (1992). A power/interaction model of interpersonal influence: French and Raven thirty years later. *Journal of Social Behavior and Personality, 7*, 217-244.

Raven, B. H. (1993). The bases of power: Origins and recent developments. *Journal of*

Social Issues, 49, 227-251.

Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Champion, K. M., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly, 23*, 10-26.

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York: Routledge.

Reinke, W. M., & Herman, K. (2002). Creating school environments that deter antisocial behaviors in youth. *Psychology in the Schools, 39,* 549-559.

Reinke, W. M., Lewis-Palmer, T., & Martin, E. (2007). The effect of visual performance feedback on teacher behavior-specific praise. *Behavior Modification, 31,* 247-263.

Reinke, W. M., Lewis-Palmer, T., & Merrell, K. (2008). The classroom check-up: A classwide consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review, 37*(3), 315-332.

Riley-Tillman, T. C., & Burns, M. K. (2009). *Single case design for measuring response to educational intervention.* New York, NY: Guilford.

Rimm-Kaufman, S. E., & Sawyer, B. E. (2004). Primary-grade teachers' self-efficacy beliefs, attitudes toward teaching, and discipline and teaching practice priorities in relation to the "Responsive Classroom" approach. *The Elementary School Journal, 104*, 321-341.

Schuldheisz, J. M., & van der Mars, H. (2001). Active supervision and students' physical activity in middle school physical education. *Journal of Teaching in Physical Education, 21*, 75-90.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York, NY: Houghton Mifflin.

Shapiro, E. S. (2004). *Academic skills problems workbook* (rev.). New York: The Guilford Press.

Sheridan, S., Welch, M., & Orme, S. (1996). Is consultation effective? A review of outcome research. *Remedial & Special Education, 17,* 341-354.

Skiba, R., Reynolds, C. R., Graham, S., Sheras, P., Conoley, J., Garcia-Vazquez, E. (2006). *Are zero tolerance policies effective in the schools? A report by the American Psychological Association Task Force* (49).

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.

Simonsen, B., Myers, D., & DeLuca, C. (2010). Teaching teachers to use prompts, opportunities to respond, and specific praise. *Teacher Education and Special Education, 33*(4), 300-318.

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*, 160-175.

Sutherland, K. S., Alder, N., & Gunter, P. L. (2003). The effect of varying rates of opportunities to respond to academic requests on the behavior of students with EBD. *Journal of Emotional and Behavioral Disorders, 11*(4), 239-248.

Sutherland, K. S., & Wehby, J. H. (2001). Relationship between increased opportunities to respond to academic requests and academic and behavioral outcomes of students with EBD: A review. *Remedial and Special Education, 22*, 113-121.

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 Behavior Disorders, 8*(1)*,* 2-8.

Theodore, L. A., Bray, M. A., Kehle, T. J. (2004). A comparative study of group contingencies and randomized reinforcers to reduce disruptive classroom behavior. *School Psychology Quarterly, 19*(3), 253-271.

Thomas, D. E., Bierman, K. L., & The Conduct Problems Prevention Research Group. (2006). The impact of classroom aggression on the development of aggressive behavior problems in children. *Development and Psychopathology, 18*, 471-487.

U.S. Department of Education, National Center for Education Statistics. (2005). *The condition of education 2005 (NCES 2005-094).* Washington, DC: U.S. Government Printing Office.

VanDerHeyden, A. M., & Witt, J. C. (2008). Best practices in can't do/won't do assessment. In A. Thomas, & J. Grimes (Eds.), *Best Practices in School Psychology* (pp. 319-336). Bethesda: The National Association of School Psychologists.

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

Wilkinson, L. A. (2006). Monitoring treatment integrity: An alternative to the 'consult and hope' strategy in school-based behavioral consultation. *School Psychology International, 29*(4), 426-438.

Witt, J. C., & Elliott, S. N. (1985). Acceptability of classroom intervention strategies. In T. R. Kratochwill (Ed.), *Advances in school psychology* (Vol. 4, pp. 251-288). Mahwah, NJ: Erlbaum.

Witt, J. C., Noell, G. H., LaFleur, L. H., & Monenson, B. P. (1997). Teacher usage of interventions in general education: Measurement and analysis of the independent variable. *Journal of Applied Behavior Analysis, 30,* 693-696.

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

Workman, E. A., & Williams, R. L. (1980). Effects of extrinsic rewards on intrinsic motivation in the classroom. *The Journal of School Psychology, 18*(2), 141-147.

Table 1

Behavioral Definitions of Classroom Management Practices

	Definition	Example	Non-Example
Prompts/Pre-correction	Specific cues that provide students with information about the behavior desired in specific situations. Prompts may be verbal, nonverbal or both.	Identifying a small number of positively stated expectations (i.e., Be Safe, Be Responsible, Be Respectful).	List of negatively stated rules (i.e., Do not talk) that do not prompt appropriate or expected behavior.
Active Supervision	The teacher moving, looking around, interacting with students, correcting errors made by students (i.e., behavior inconsistent with expectations), and providing reinforcement for behavior consistent with expectations.	Teacher is circulating around the room, scanning the classroom, speaking to students, and providing reinforcement for meeting expectations (i.e., "Suzy, I like the way you are raising your hand to speak").	Teacher sitting at his or her desk, looking away from students, and not providing reinforcement for desired behaviors.
Opportunities to respond	Teacher behavior that prompts or solicits a student response.	Erasable boards on which students write their answers to a question. Teacher cues class to respond chorally.	Students shouting answers without being asked to.
Specific praise	Positive statement provided by the teacher when a desired behavior occurs to inform	Teacher saying, "Bobby, I like the way you raised your hand to speak."	Teacher saying, "Good job."

	students specifically what they did well.		
Error correction	Informative statement provided by the teacher given when an undesired behavior occurs and tells the student exactly what they should do in the future, in a brief, concise manner.	Teacher saying, “Bobby, please keep your bottom on the chair and feet on the floor.”	Teacher saying, “Bobby, sit correctly.”

Table 2

Study Components

Component	Description
Procedures	Teachers consent to participate Teacher and consultant meet to discuss process Baseline begins and observation data collected 3 times a week After baseline, weekly performance feedback Process continues for 10 weeks
Baseline	Observation data collected 3 times a week Weekly meetings for 5-10 minutes Collect teacher survey and student demographic data Scheduling and discussion of study procedures
Consultation Process	Share data Discuss antecedents and consequences Develop Plan
Performance Feedback	5-10 minutes Present classroom management and student engagement data on graph Review data, provide specific praise and problem solving to improve implementation Schedule next meeting
Problem Evaluation	Final performance feedback meeting Review final data Examine plan effectiveness Ensure maintenance of effect Administer social validity survey

Table 3

Teacher and Classroom Demographic Data

ID	Age	Gender	Ethnicity	Grade	Years Teaching	Admin Support	PD	Behavior Classes	ODRs
T1	4	F	3	K	26	3	1-2	1-2	6
T2	3	F	1	3	27	3	1-2	1-2	4
T3	3	F	2	6	26	4	1-2	4-5	8
T4	4	M	4	5	13	6	1-2	1-2	12
T5	3	M	3	3	1	6	1-2	1-2	1

Note. Age: 0 = 22-25, 1 = 26-30, 2 = 31-40, 3 = 41-50, 4 = 51-60, 5 = 60 or over; Ethnicity: 0 = American Indian/Alaska Native, 1 = Asian American/Pacific Islander, 2 = Black/African American, 3 = Caucasian, 4 = Hispanic; Administrator Support: 1 = Strongly Disagree, 2 = Disagree, 3 = Disagree Somewhat, 4 = Agree Somewhat, 5 = Agree, 6 = Strongly Agree.

Table 4

Effect Size Statistics

	PAND	Phi	Interpretation (Cohen, 1988)
Prompts	55.0%	.10	Small
Active Supervision	65.9%	.31	Medium
OTRs	52.7%	.05	Small
Praise	95.3%	.91	Large
Error Correction	84.5%	.69	Large
Total Engagement	98.5%	.97	Large
Active	65.2%	.30	Medium
Passive	65.9%	.31	Medium
Disruptive	83.7%	.67	Large

Note. PAND = Percent of All Non-overlapping Data; OTRs = Opportunities to Respond.

Table 5

Social Validity Survey

	Total	*M*	*SD*
Teacher 1	90	6	0
Teacher 2	71	4.7	0.7
Teacher 3	79	5.3	1.1
Teacher 4	90	6	0
Teacher 5	90	6	0

Note. M = Mean; *SD* = Standard Deviation.

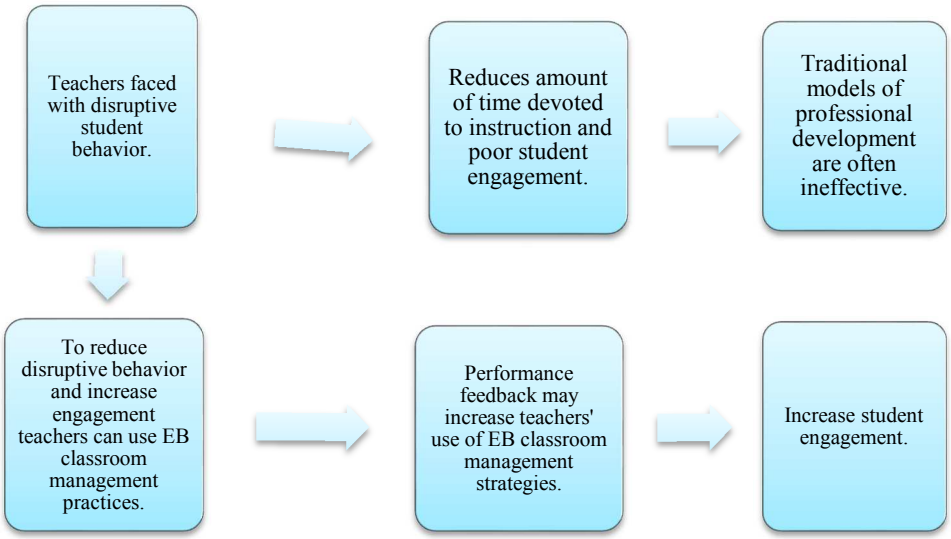

Figure 1. Logic model presenting a graphical depiction of the relationship between providing performance feedback and teacher and student outcomes.

Teacher Survey

1. Grade level you are now teaching:
2. Gender:
   Male Female
3. Age:

22-25 26-30 31-40 41-50 51-60 60 or over

4. Ethnicity:
American Indian/Alaska Native Asian American/Pacific Islander
Black/African American Caucasian Hispanic Other Do Not Want to Specify

4. Years of experience:

5. Highest degree earned:

BA/BS MS/MA Specialist Doctorate

6. Amount of training in behavior management:

1-2 classes 3-4 classes 4-5 classes 6 or more classes

7. How many professional developments addressing behavior management have you received through the district?

0 1-2 3-4 5-6 7 or more

8. Do you feel your administrator provides support regarding behavior concerns?

1	2	3	4	5	6
Strongly Disagree	Disagree	Disagree Somewhat	Agree Somewhat	Agree	Strongly Agree

Figure 2. Teacher demographic survey utilized during weekly baseline meetings to inform description of participants.

Classroom Observation of Teachers and Students (COTS)

Date: Teacher:

Time: Subject: Observer:

Moment	1	2	3	4	5	6	7	8	9	10		Total
Prompts/Pre-corrections											AET	
Active Supervision											PET	
Providing OTRs											OT	
Specific Praise												
Error Correction												

Moment	11	12	13	14	15	16	17	18	19	20		Total
Prompts/Pre-corrections											AET	
Active Supervision											PET	
Providing OTRs											OT	
Specific Praise												
Error Correction												

Moment	21	22	23	24	25	26	27	28	29	30		Total
Prompts/Pre-corrections											AET	
Active Supervison											PET	
Providing OTRs											OT	
Specific Praise												
Error Correction												

Moment	31	32	33	34	35	36	37	38	39	40		Total
Prompts/Pre-corrections											AET	
Active Supervison											PET	
Providing OTRs											OT	
Specific Praise												
Error Correction												

Moment	41	42	43	44	45	46	47	48	49	50		Total
Prompts/Pre-corrections											AET	
Active Supervison											PET	
Providing OTRs											OT	
Specific Praise												
Error Correction												

Moment	51	52	53	54	55	56	57	58	59	60		Total
Prompts/Pre-corrections											AET	
Active Supervison											PET	
Providing OTRs											OT	
Specific Praise												
Error Correction												

	T	O	%
Prompts/Pre-Corrections			
Active Supervison			
Providing OTRs			
Specific Praise			
Error Correction			

	T	O	%
AET			
PET			
OT			

Setting: ISW: TPsnt SmGp:TPsnt

ISW: TSmGp LgGp:TPsnt

Other:

Figure 3. Direct observation measure of classroom management practices and student engagement. Adapted from "Behavioral Observation of Students in Schools (BOSS), by E. S. Shapiro, *Academic skills problems workbook* (rev.). Copyright 2004 The Guilford Press.

Consultation Procedural Checklist

Consultant: Evaluator: Date:

Components	Observed	Not Observed	Comments
Problem Identification: Identifies the problem in operational terms			
Identifies the settings and causes			
Identifies behavior strength			
Potential goals established			
Problem Analysis: Data collected in baseline is reviewed			
Antecedents, consequences, and patterns are discussed			
Plan is made			
Performance Feedback: 5-10 minutes			
Receives graphed teacher and student behavioral data			
Data reviewed, praise provided for completed steps & problem solving for improving implementation			
Plan Evaluation: Review final data			
Evaluate goal attainment			
Questions about			

goal attainment			
Evaluate plan effectiveness			
Evaluate external validity of plan			
Post implementation planning occurs			
Plan modification if needed			
Procedures to facilitate generalization & maintenance/follow up assessment			
Termination of consultation			
Social validity measure given			

Figure 4. Fidelity checklist utilized after consultation sessions to determine if the consultation procedures were implemented as designed. Items pertain to the consultation meetings and weekly performance feedback meetings.

Performance Feedback Procedures Rating

1. The performance feedback procedures used would be acceptable for our school.

1	2	3	4	5	6
Strongly Disagree	Disagree	Disagree Somewhat	Agree Somewhat	Agree	Strongly Agree

2. Most teachers would find performance feedback appropriate.

1	2	3	4	5	6
Strongly Disagree	Disagree	Disagree Somewhat	Agree Somewhat	Agree	Strongly Agree

3. Performance feedback should prove effective in meeting the purposes.

1	2	3	4	5	6
Strongly Disagree	Disagree	Disagree Somewhat	Agree Somewhat	Agree	Strongly Agree

4. I would suggest the use of performance feedback to other teachers.

1	2	3	4	5	6
Strongly Disagree	Disagree	Disagree Somewhat	Agree Somewhat	Agree	Strongly Agree

5. Performance feedback is appropriate to meet the school's needs.

1	2	3	4	5	6
Strongly Disagree	Disagree	Disagree Somewhat	Agree Somewhat	Agree	Strongly Agree

6. Most teachers would find performance feedback suitable for the described purposes.

1	2	3	4	5	6

Strongly Disagree	Disagree	Disagree Somewhat	Agree Somewhat	Agree	Strongly Agree

7. I would be willing to use performance feedback in the school setting.

1	2	3	4	5	6
Strongly Disagree	Disagree	Disagree Somewhat	Agree Somewhat	Agree	Strongly Agree

8. Performance feedback would *not* result in negative side effects for the students.

1	2	3	4	5	6
Strongly Disagree	Disagree	Disagree Somewhat	Agree Somewhat	Agree	Strongly Agree

9. Performance feedback would be appropriate for a variety of students.

1	2	3	4	5	6
Strongly Disagree	Disagree	Disagree Somewhat	Agree Somewhat	Agree	Strongly Agree

10. Performance feedback is consistent with those I have used in school settings.

1	2	3	4	5	6
Strongly Disagree	Disagree	Disagree Somewhat	Agree Somewhat	Agree	Strongly Agree

11. Performance feedback is a fair way to fulfill the strategies purposes.

1	2	3	4	5	6
Strongly Disagree	Disagree	Disagree Somewhat	Agree Somewhat	Agree	Strongly Agree

12. Performance feedback is reasonable to meet the stated purposes.

1	2	3	4	5	6
Strongly Disagree	Disagree	Disagree Somewhat	Agree Somewhat	Agree	Strongly Agree

13. I like the procedures used in performance feedback.

1	2	3	4	5	6
Strongly Disagree	Disagree	Disagree Somewhat	Agree Somewhat	Agree	Strongly Agree

14. Performance feedback is a good way to meet the specified purpose (increase student engagement).

1	2	3	4	5	6
Strongly Disagree	Disagree	Disagree Somewhat	Agree Somewhat	Agree	Strongly Agree

15. Overall, performance feedback would be beneficial for elementary school students.

1	2	3	4	5	6
Strongly Disagree	Disagree	Disagree Somewhat	Agree Somewhat	Agree	Strongly Agree

Figure 5. Social validity questionnaire used to measure teacher perceptions of the social significance of intervention goals, social acceptability of intervention procedures, and likelihood of socially important outcomes. Adapted from "Acceptability of Classroom Intervention Strategies," by J. C., Witt, and S. N. Elliott, 1985, *Advances in School Psychology,* 4, p. 251. Copyright 1985 by Erlbaum, Mahwah, NJ.

Problem Identification Interview

Consultant: _____
Teacher name: _____
Grade: _____
Date of Assessment: _____

Intro: What is the problem?

- How do you define classroom management?
- What are some classroom management practices you use?
- What are your classroom rules?
- How did you decide on these classroom rules?
- What are your classroom consequences?
- How do you handle students who chronically misbehave?
- What are the behaviors you are concerned about today?
- Can you describe for me your students' general classroom behavior?
- Can you give me a recent example of what this problem behavior looks like as well as what it is not?
- What do students do when they are _____?
- Are there any other examples you can think of that really give me a good picture of these behaviors in the classroom?
- Of the behaviors ___, _____, & _____, which would you consider to be the most problematic and of concern to you.
    - How severe is the problem for you?

*So, you've said that the current concern about _____ is when they _____. Do I have that information right?

Setting & causes:

- During the school day, when do these problem behaviors generally occur?
    - Is it primarily a certain time of day or in one particular class/lesson?
- Are there other times when the behaviors occur?

    - Other classes, lessons, teachers?
- Of the different places and times that this problem is of concern, when is it the most difficult?
- What happens right before these problem behaviors occur?
    - What happens right before the students do "behavior"?
- What is the general pattern of the behavior and what happens prior to it across several occurrences of the behavior?
    - Who are the students with?
    - What are the students supposed to be doing?
- What happens after the students do "behavior"?

- What is your response and actions after _____?

- What do other children do when students do _____?

*So, you've said that the concern with your students is when they _____ and that this behavior generally follows_____ and _____ happens after. Is this information correct?

Behavior strength:

- How often do _____ occur?
    - Weekly, daily?

- How long do these behaviors generally last for?

*So, the behaviors are a problem _____ # of times for _____ length?

Potential goal:

- What is a reasonable amount of behavior reduction you could expect from your students to consider the behavior to have improved?
    - How frequently would _____ have to not interrupt for it to not cause a problem?

- During instruction, how long are students doing seatwork?

*Let's see, the main problem is that students do "behavior" during _____ about _____ days a week/times a day. Is all of this right?

- When is a good time for me to come back and do our follow-up?

Problem Analysis Interview

- Let's take a look at the record that's been collected on your students' behavior
  - General observation about data and problem

- It looks as if your students did _____
  - Questions or statements about the strength of behavior

- From my notes, I noticed some things that happened just before or after the behavior occurred.
  - Antecedents
  - Consequences
  - Pattern of sequences during behavior

- Let's see, so the students' was "behavior" by doing _____ last week __# of times. This behavior seemed to be related to "antecedent". We would like to see this behavior eliminated to help _____. Is this right?
  - Target what behavior is, the conditions and strength

- Why do you think your students have been _____?
  - Interpret the behavior

- We need to try something different with your students. What could be done before your students engage in the "behavior"? How could we remove the attention from the behavior?
  - Questions about the plan

- Then we will go ahead and try _____
  - Validate the plan

Treatment Evaluation Interview

- Opening salutation
- Evaluate goal attainment: Questions about outcome
  - How did things go?
- Questions about goal attainment
  - Are students more on-task/less disruptive?
  - Can we say that the goal of increasing students' on-task behavior has been attained now?
- Evaluate plan effectiveness: Questions regarding internal validity of plan.
  - Would you say that the classroom management practices were responsible for reducing disruptive student behavior and increasing on-task behavior?
- Evaluate external validity of plan
  - Do you think this plan would work with another classroom?
- Conduct post implementation planning: Questions and statements regarding plan continuation.
  - Do you want to leave the practices in place for another week to see if progress continues?
  - Or perhaps we should try another classroom management strategy?
- Questions and statements regarding plan modification
  - You are saying you want to discontinue the plan….
  - How could we change the procedure to make our plan more effective?
  - Perhaps we could…

- Design procedures to facilitate generalization and maintenance
    - What procedures can be implemented to be sure that students' continue to stay on-task?
- Arrange for follow-up assessment.
    - Now that we have success in the program for your students, how can we monitor their progress in the future?
- Arrange for subsequent interviews or terminate consultation: Questions and statements regarding future interviews.
    - When can we get together again to discuss your students' progress under our new plan?
    - We probably need to meet again next week to discuss our new plan.
- Statements regarding termination of consultation.
    - Since our goals have been met, this will be the last time we meet unless you have further concerns.
    - If you have further concerns, please feel free to contact me.
- Closing salutation

Figure 6. Class-wide versions of the Problem Identification Interview, Problem Analysis Interview, and Treatment Evaluation Interview for classroom management behavior issues. Adapted from "Behavioral Consultation in Applied Settings: An Individual Guide" by T. R. Kratochwill and J. R. Bergan, 1990. Copyright 1990 by Plenum Press: New York.

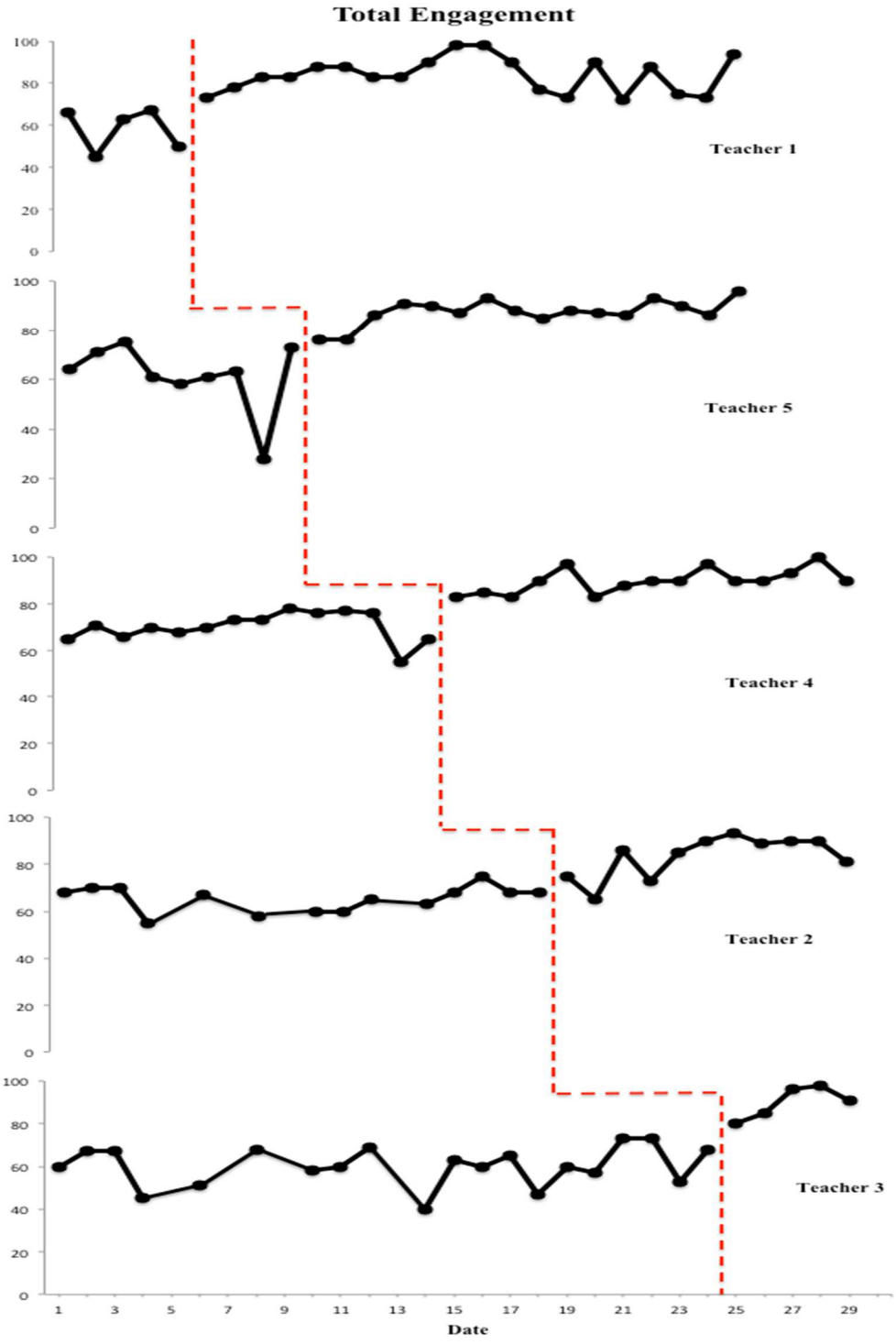

Figure 7a. Percent of intervals of total student engagement during time observed.

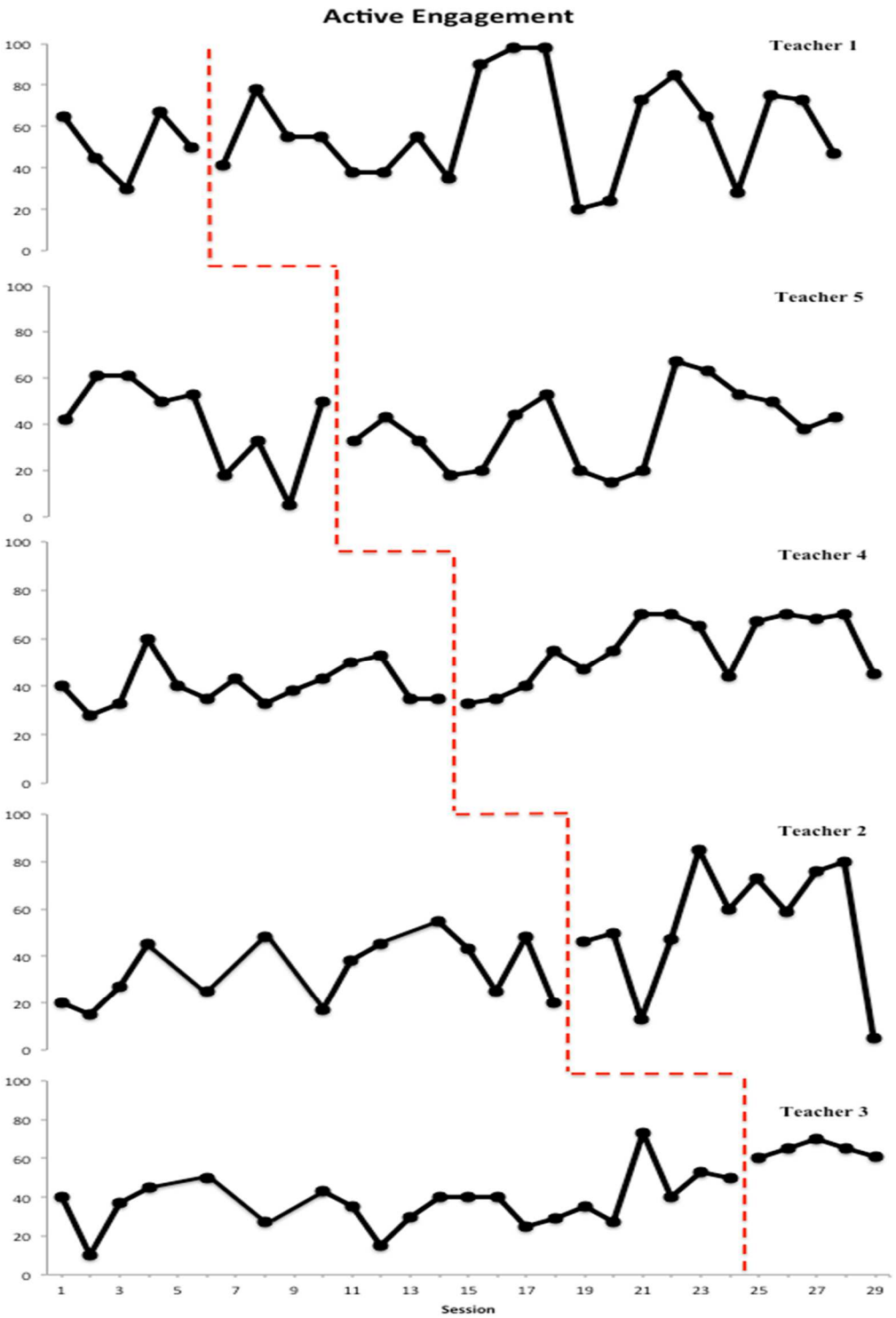

Figure 7b. Percent of intervals of active engagement during time observed.

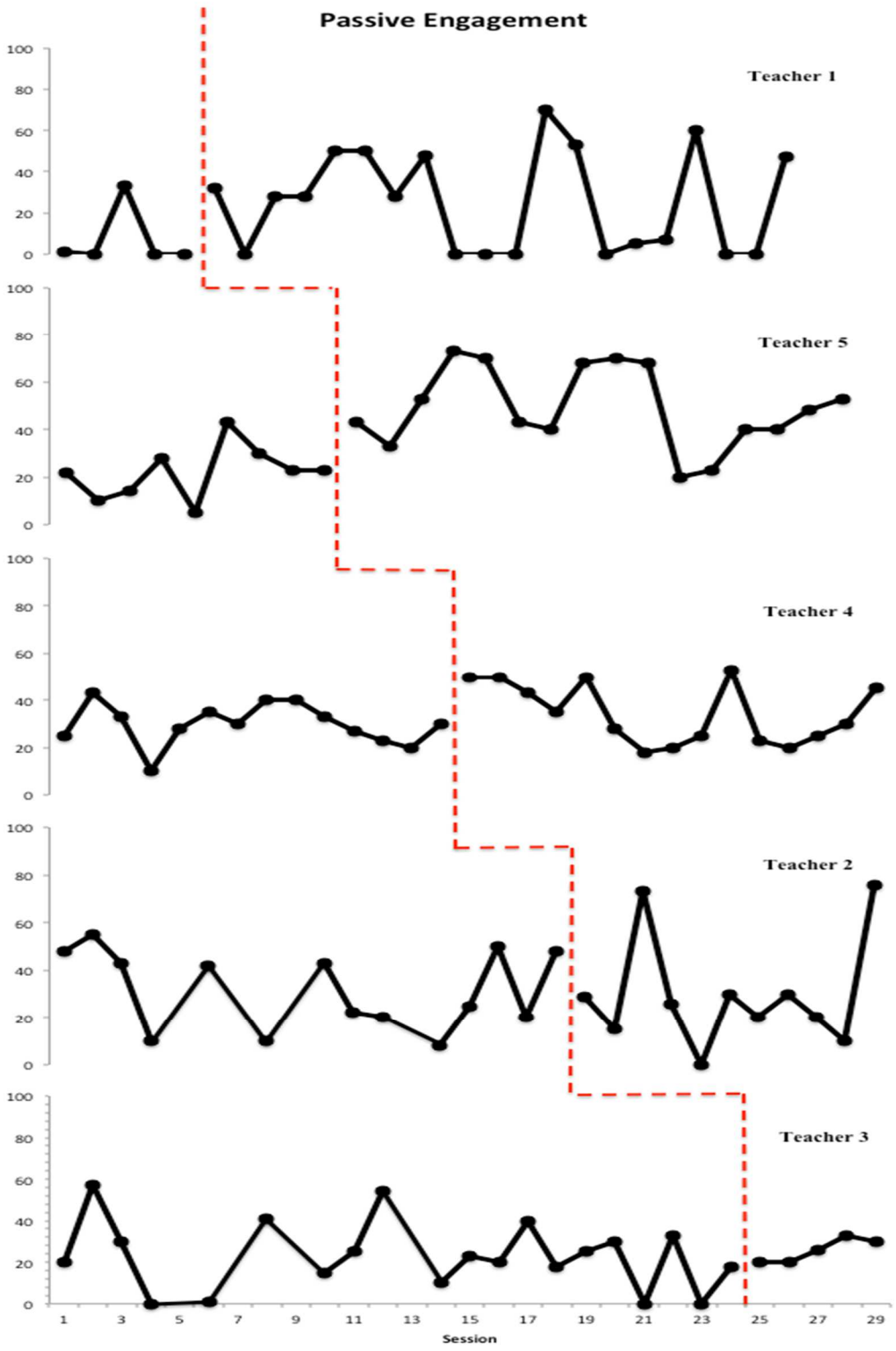

Figure 7c. Percent of intervals of passive engagement during time observed.

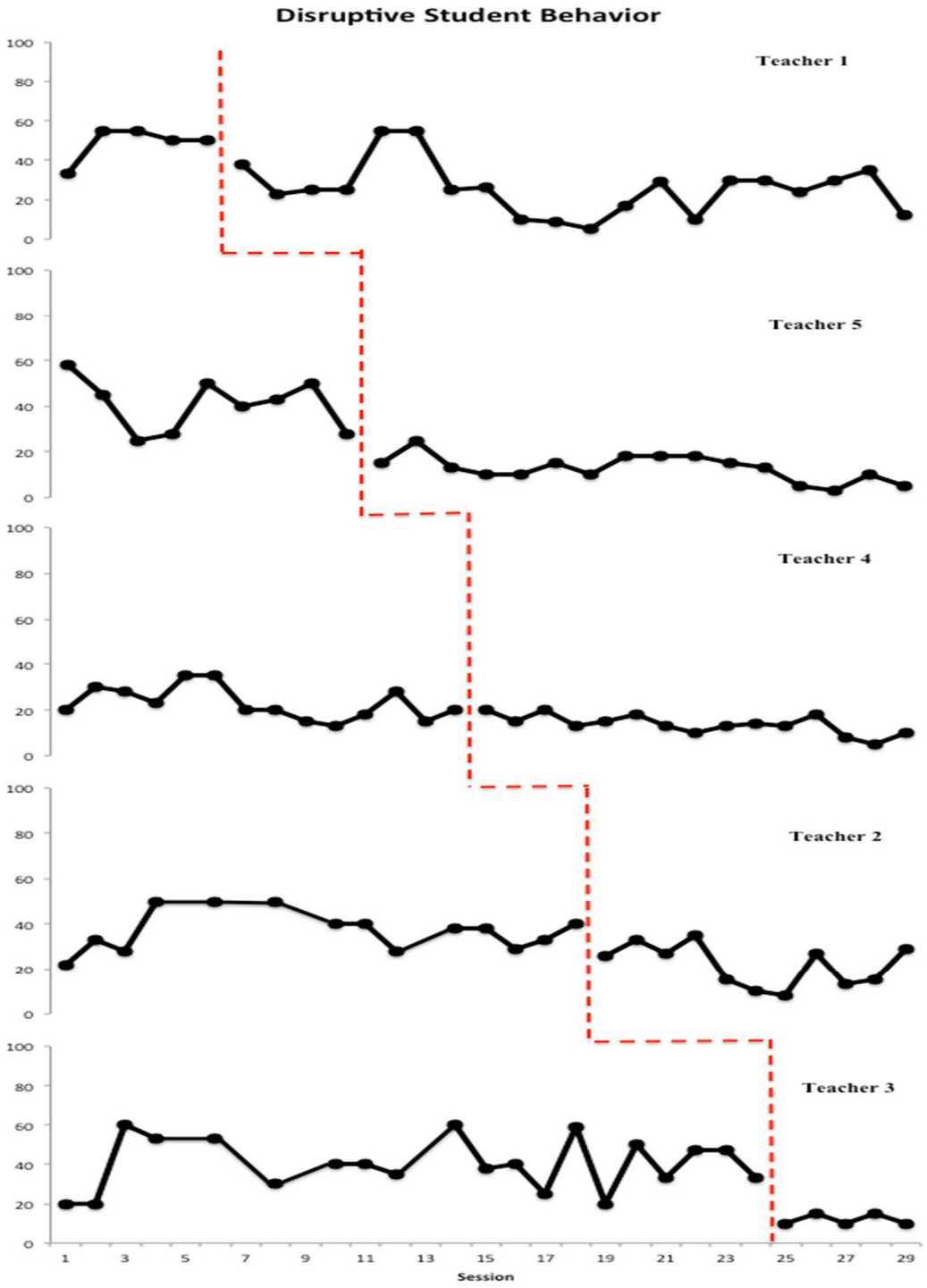

Figure 7d. Percent of intervals of disruptive student behavior during time observed.

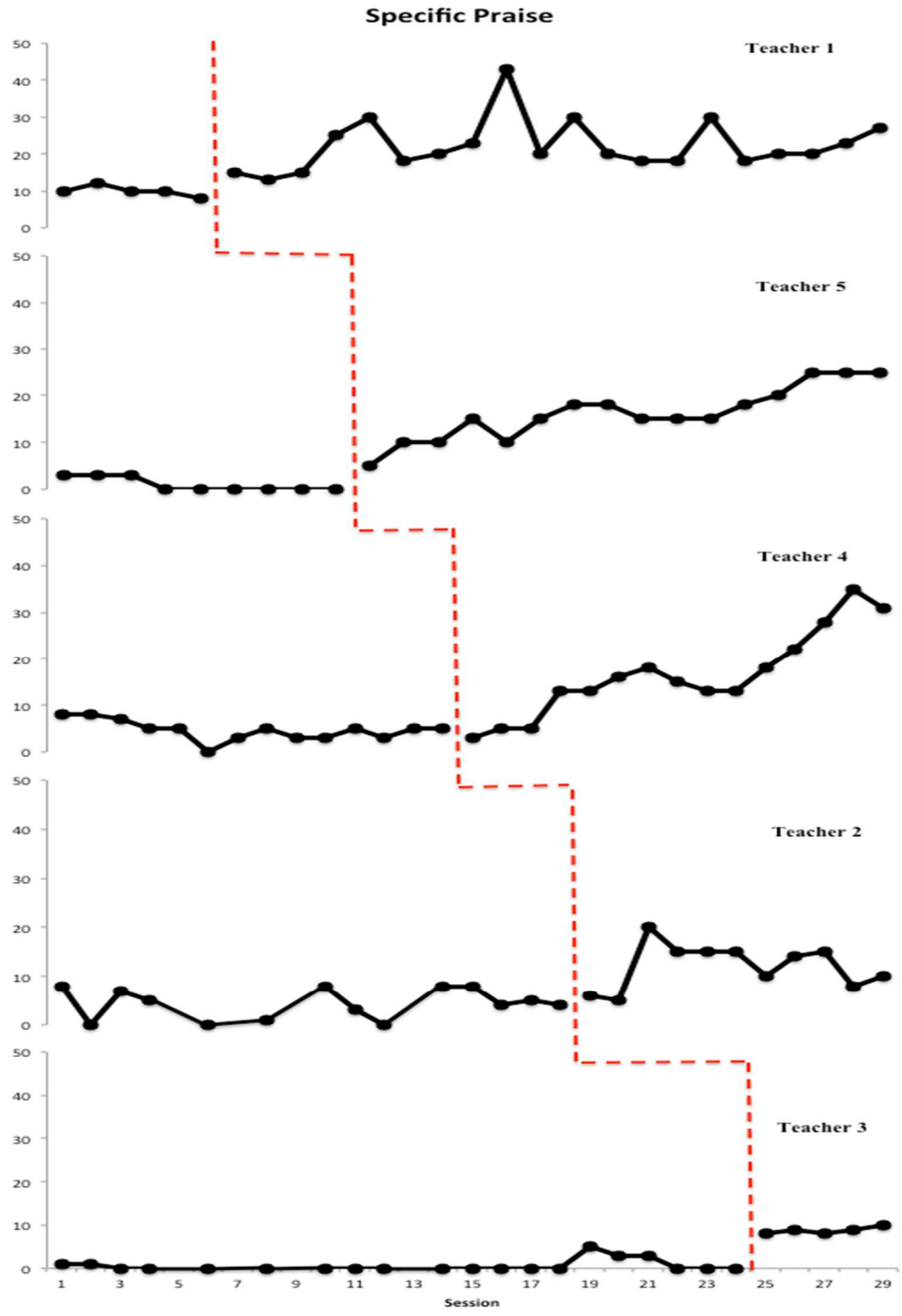

Figure 7e. Percent of intervals of specific praise during time observed.

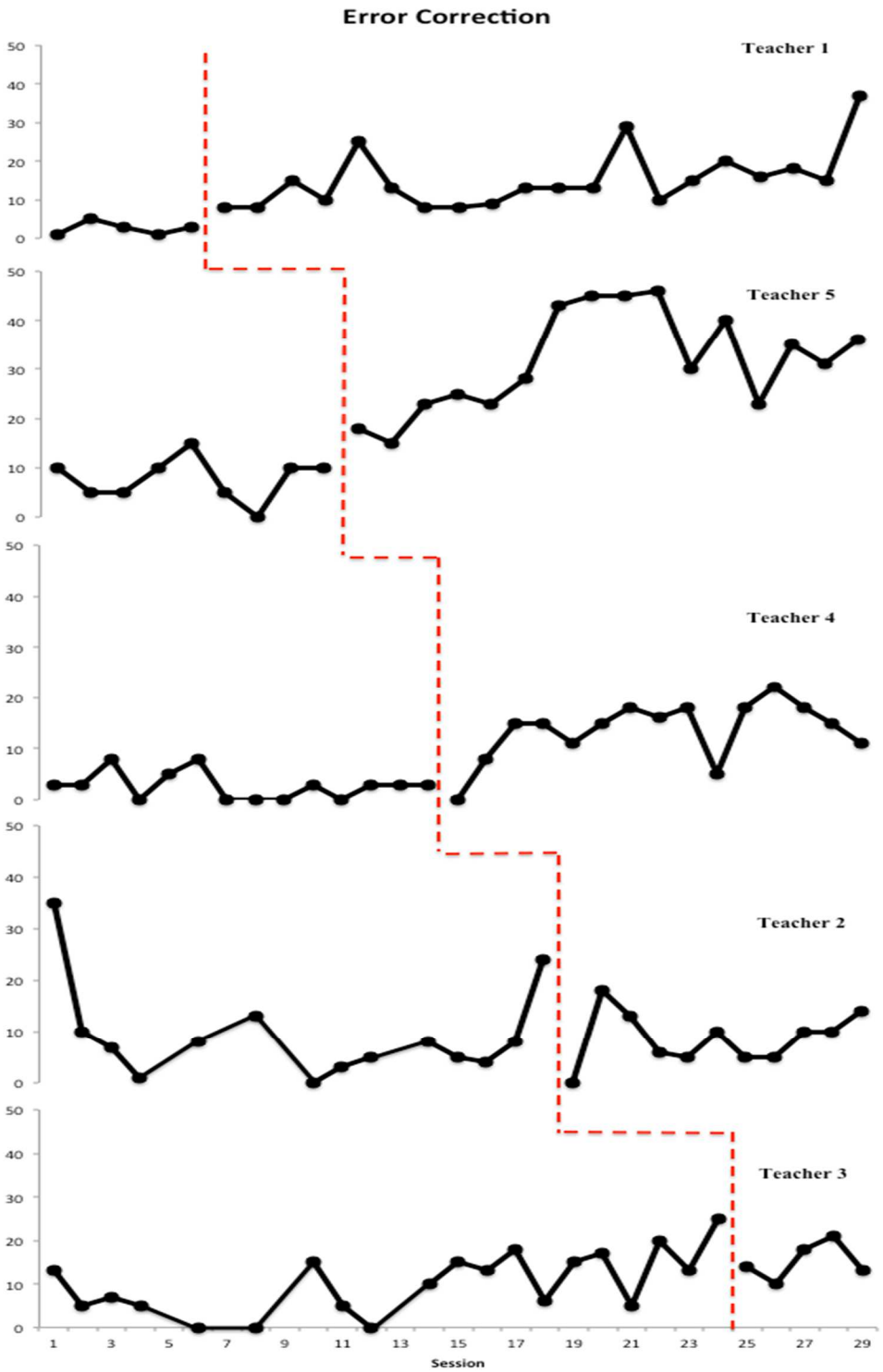

Figure 7f. Percent of intervals of error corrections during time observed.

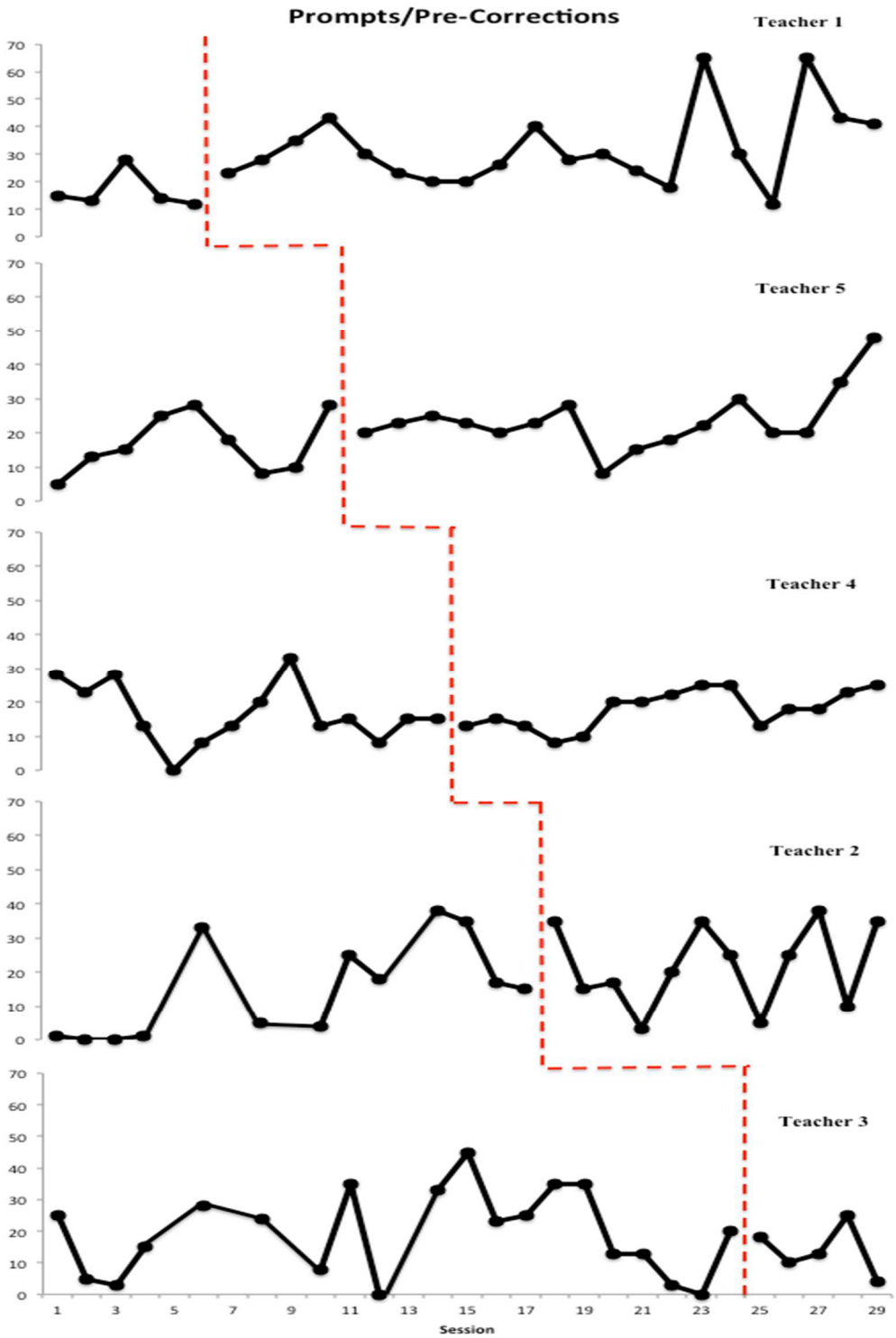

Figure 7g. Percent of intervals of prompts/pre-corrections during time observed.

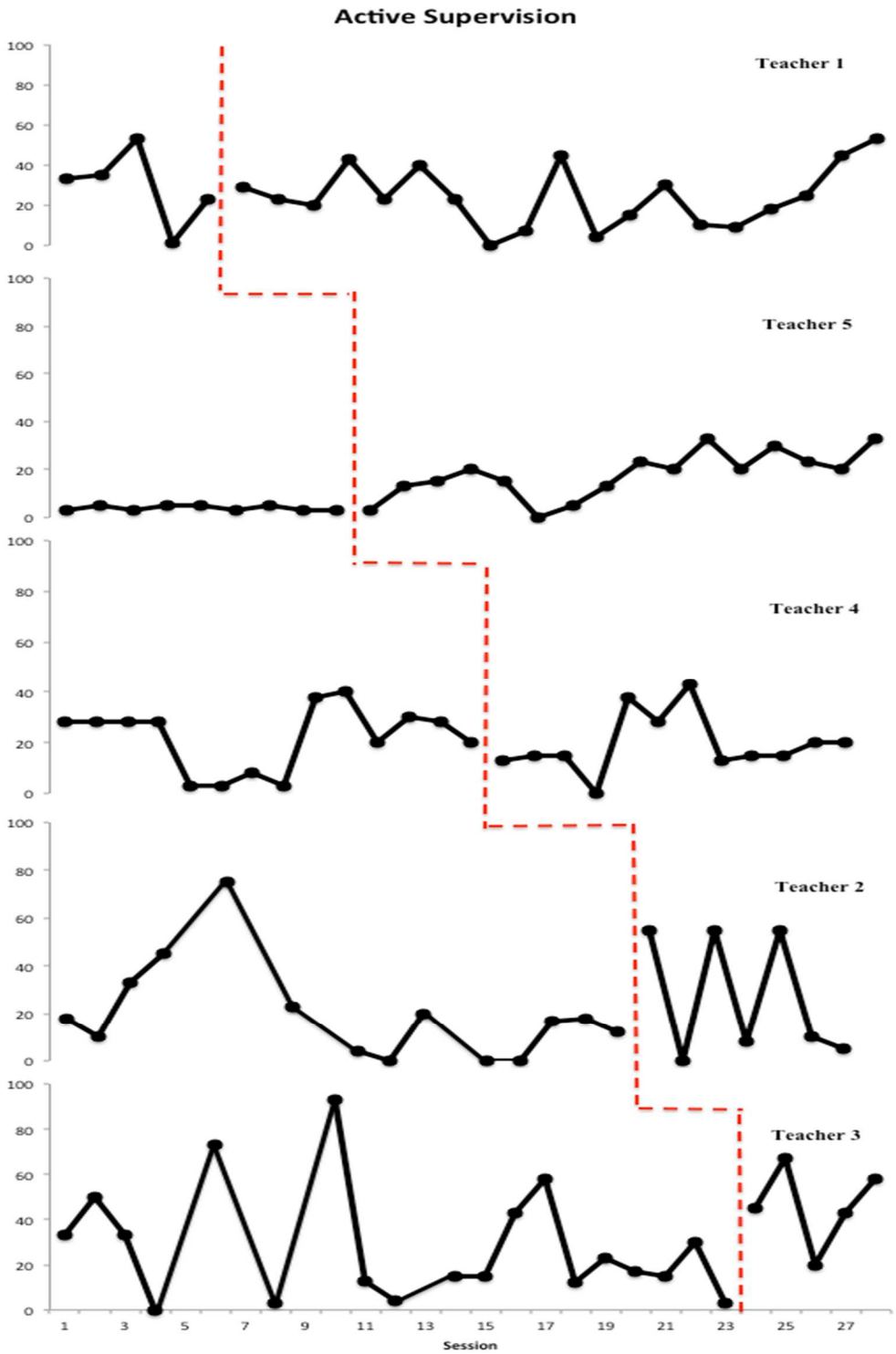

Figure 7h. Percent of intervals of active supervision during time observed.

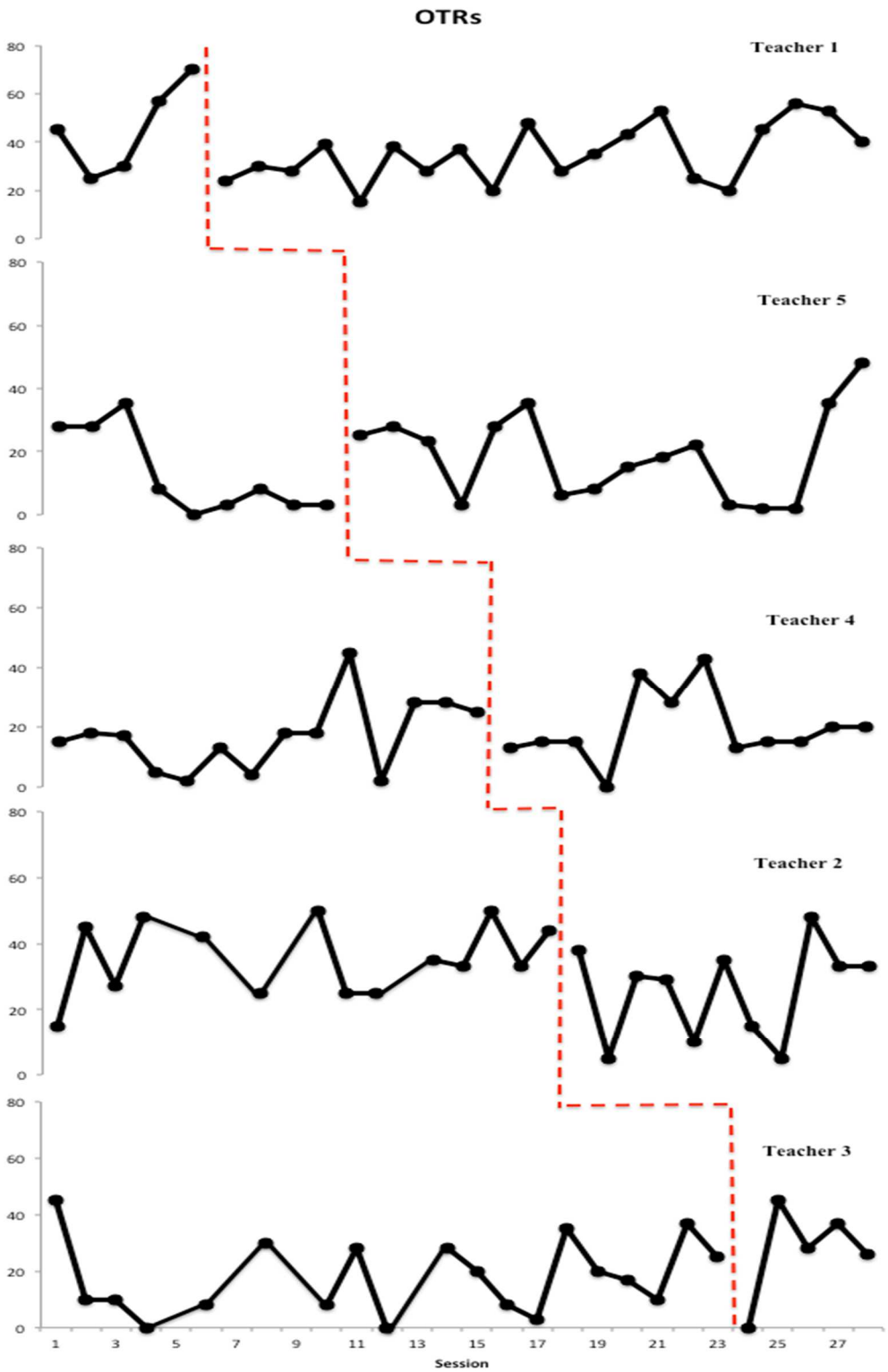

Figure 7i. Percent of intervals of opportunities to respond during time observed.