

Evaluating the Effects of Performance Feedback Using the Direct Behavior Rating-
Classroom Management (DBR-CM) on Teacher Classroom Management Behavior

A Dissertation presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by
Wesley A. Sims
Dr. T. Chris Riley-Tillman, Dissertation Supervisor
July, 2016

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

EVALUATING THE EFFECTS OF PERFORMANCE FEEDBACK USING THE
DIRECT BEHAVIOR RATING-CLASSROOM MANAGEMENT (DBR-CM) ON
TEACHER CLASSROOM MANAGEMENT PERFORMANCE

presented by Wesley A. Sims, a candidate for the degree of doctor of philosophy, and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Chris Riley-Tillman, Ph.D.

Professor Wendy M. Renike, Ph.D.

Professor Erica Lembke, Ph.D.

Professor Aaron Thompson, Ph.D.

Professor Stephen Kilgus, Ph.D.

ACKNOWLEDGEMENTS

Over the past four years I have received support and encouragement from a great number of individuals. Dr. T. Chris Riley-Tillman has been a mentor, colleague, and friend. His guidance has made this a thoughtful and rewarding journey. In addition, Drs. Wendy Reinke and Stephen Kilgus were instrumental to my successful completion of this project. I would like to thank the remaining members of my dissertation committee, Drs. Erica Lembke and Aaron Thompson for their support over the past two years. I also wish to thank Dr. Keith Herman. Though not formally connected to this project, he was instrumental to the organization and development of the manuscript for this project.

I cannot thank both the participants and data collectors involved in this project enough for their contributions. Their time, effort, and energy all but donated to this project is sincerely appreciated. A special thanks goes to Dr. Katie King for the countless hours spent proofreading and listening to me talk about this and my other research projects.

TABLE OF CONTENTS

Acknowledgements	ii
List of Tables	vi
List of Figures	vii
Abstract.....	viii
Chapter 1	
Introduction.....	1
Chapter 2	
Role of Teacher in Learning.....	10
Classroom Management Importance	12
Training in Classroom Management.....	13
Professional Development Activities and Classroom Management.....	15
Treatment Integrity and Classroom Management.....	17
Performance Feedback.....	19
Difficulties with Performance Feedback.....	20
Advances in Performance Feedback.....	21
The Classroom Check Up	22
Assessment for Performance Feedback	25
Systematic Direct Observation.....	27
Classroom Management Assessment.....	29
Alternative Classroom Management Assessment.....	31
Direct Behavior Ratings Background.....	32
General Outcomes Measures Background.....	33
Direct Behavior Ratings-Single Item Scales Background.....	34
Direct Behavior Ratings-Classroom Management Development.....	35
Research Questions.....	40
Chapter 3	
Method.....	42
Participants.....	42
Teacher 1.....	42
Teacher 2.....	42
Teacher 3.....	42
Teacher 4.....	43
Teacher 5.....	43
Measures.....	43
Brief Classroom Interaction Observation-Revised	43
Direct Behavior Rating-Classroom Management.....	44
User Rating Profile-Intervention.....	46

Procedures	46
Dependent Variable.....	48
Independent Variable.....	48
Design.....	51
Single Case Design.....	51
Multiple-Baseline Design.....	53
Analytic Plan.....	56
Evaluating Results.....	59
Chapter 4	
Results	61
Fidelity Indices	61
Visual Analyses.....	61
Teacher 1.....	61
Teacher 2.....	62
Teacher 3.....	62
Teacher 4.....	62
Teacher 5.....	63
Empirical Analysis.....	63
Social Validity.....	65
Chapter 5	
Discussion	67
Limitations	74
Future Directions.....	75
Conclusion.....	78
References	80
Table 1.....	95
Table 2.....	97
Table 3.....	98
Figure 1	100
Figure 2.....	101
Table 4.....	102
Table 5.....	103
Table 6.....	104
Table 7.....	105

Figure 3106

Vita107

LIST OF TABLES

Table	Page
1. Operational Definitions of Direct Behavior Rating-Classroom Management	95
2. Pertinent Participant Demographic Data	97
3. Operational Definitions of Direct Observation Variables	98
4. Slope Across Phases for BCIO-R: Overall Rate of Praise	102
5. Advanced Empirical Analyses of Intervention Effects on Rate of Praise	103
6. BCIO-R: Overall Rate of Praise Contingency Table Analysis	104
7. User Rating Profile-Intervention Mean Item Score	105

LIST OF FIGURES

Figure	Page
1. Example of Graphic Representation of Performance Feedback Provided to Participating Teachers	100
2. BCIO-R: Overall Rate of Praise	101
3. Visual Presentation of BCIO-R: Overall Rate of Praise and DBR-CM Praise t-scores	106

EVALUATING THE EFFECTS OF PERFORMANCE FEEDBACK USING THE
DIRECT BEHAVIOR RATING - CLASSROOM MANAGEMENT (DBR - CM) ON
TEACHER CLASSROOM MANAGEMENT BEHAVIOR

Wesley A. Sims

T. Chris Riely-Tillman – Dissertation Supervisor

ABSTRACT

Despite the importance of effective classroom management, teachers often report receive surprisingly little classroom management support. Performance feedback has been used to further develop teacher classroom management behavior. To supplement available classroom management assessments with which to provide performance feedback, The Direct Behavior Rating-Classroom Management (DBR-CM) was developed. The present study used a multiple baseline single-case design research design to examine a simplified performance feedback intervention targeting teacher classroom management behavior. Performance feedback was delivered through graphic representation of DBR-CM assessment results following external observation to participating teachers displaying mild to moderate deficits in classroom management performance. Visual and empirical analyses of results indicated noteworthy improvements in overall teacher classroom management behavior in two of five included teachers. These findings indicate a performance feedback classroom management intervention in this form failed to document the required three replication of intervention effects to be considered efficacious. Several factors may explain these findings including study timing, inappropriate inclusion of participants, teacher resistance, shortcomings in performance feedback information provided, and shortcomings of performance feedback dissemination methods.

Chapter 1

Introduction

Research has demonstrated that children in well-managed classrooms spend more time engaged in academic tasks, progress at a more rapid pace, and demonstrate higher levels of academic achievement (Brophy, 1985). Effective teacher classroom management behaviors have been shown to predict student engagement and social competence, as well as reduce disruptive behaviors. La Paro & Pianta (2003) note well managed, optimally performing classrooms are characterized by low levels of conflict and disruptive behavior, smooth transitions from one type of activity to another, appropriate expressions of emotion, respectful communication and problem solving, strong interest and focus on task, and supportiveness and responsiveness to individual differences and students' needs. Well-managed classrooms are distinguished by teachers' ability to monitor student attention and performance, establish behavioral expectations, and consistently implement rules and procedures that prevent problems from occurring (Brophy, 1985; Reinke, Lewis-Palmer, & Merrell, 2008). Alternatively, a poorly managed classroom, or a classroom in which a teacher(s) lacks the skill or resources to effectively manage the challenges within the particular context of their school and classroom, results in lower levels of on-task behavior and performance (Brophy, 1985; Reinke et al., 2008).

Despite the importance of effective classroom management, teachers typically receive surprisingly little classroom management support (Wagner et al., 2006). Upon completion of their formal university training and student teaching experience, little meaningful effort is devoted to further developing the performance of teachers (Pindiprolu, Peterson, & Bergloff, 2007). A frequent challenge faced by trainers of

teachers is in the identification of effective and relevant ways to improve teacher behavior when current levels of performance are not optimal (Myers, Simonsen, & Sugai, 2011). The “in-service approach” to professional development is often ineffective, and typically more informational than developmental (Guskey, 2000). Training is typically delivered in relatively passive, one-session workshops with few opportunities to achieve skill fluency through practice or feedback (Timperley, 2008). Thus, these types of events rarely result in durable change in teacher practices (Guskey, 2000). To address the limitations of this approach, professional development efforts are shifting from “one-shot” style training events to more systematic and continuous teacher support practices (Sugai et al., 2000). One prominent attempt to intervene in classroom management that has gained scholarly attention is seen in the use of performance feedback within a school-based consultation (SBC) framework (Erchul & Ward, 2016; Reinke, Herman, & Sprick, 2011).

SBC, which has been a frequently used method to address student academic and behavioral difficulties (Bergan, 1977), shows promise in addressing classroom management deficiencies. Within most SBC approaches, a persistent limitation is the attribution of problematic functioning to within individual student(s) factors (Sheridan, Welch, & Orme, 1996). This attribution places the focus of the consultation relationship on individual student problems (academic or behavioral) or a small group of students, rather than on changing the classroom system (Reinke et al., 2008). This results in intervention efforts that target students rather than a change in the entire classroom ecology, or classroom management behaviors. Development of intervention strategies targeting the improvement of classroom management behavior explicitly represents an

area in the school setting that could dramatically impact a large number of students (Reinke, Herman, & Stormont, 2013).

McKinney and colleagues (1983) identified performance feedback within SBC as an opportunity for increasing teacher skills. Recent SBC literature focuses on the use of performance feedback as a method of increasing the treatment integrity of prescribed interventions (Solomon, Klein, & Politylo, 2012), and as an effective method of teacher behavior change (Reinke et al., 2014; Reinke, Stormont, Webster-Stratton, Newcomer, & Herman, 2012). Performance feedback, or providing ongoing, data-based, objective feedback on the current performance of targeted teacher behaviors, is among the most promising methods for increasing the implementation of evidence-based instructional practices (Reinke, Lewis-Palmer, & Martin, 2007; Sugai et al., 2000). Three prominent forms of performance feedback for teachers in the school setting are evaluation of treatment integrity for a specific intervention with a child or small group, summative evaluations conducted by building or district administrators, and direct feedback as part of the consultation process (Noell, Witt, Gilbertson, Ranier, & Freeland, 1997; Sanetti, Luiselli, & Handler, 2007; Scheeler, Ruhl, & McAfee, 2004; Showers, 1985).

Despite the promise shown by performance feedback, the attribution of student difficulty to a “within-student” issue continues to be a limiting factor within its application. More modern models of consultation attempt to shift the emphasis from students to the instructional environment as the target of assessment and intervention (Reinke et al., 2008). This approach views a student as part of the instructional system rather than the primary source of the problem (Reinke et al., 2008). Such approaches emphasize the influence environmental factors play in contributing to problematic

functioning, and a prevention, early intervention, and problem solving approach to addressing the impact of these factors.

Teacher resistance to a SBC may pose a barrier to effective implementation of the CCU, as consultative relationships may be interpreted as an overt indication of teacher performance deficits (Gutkin & Hickman, 1990). In cases where teachers do not solicit consultation, engaging in a consultative relationship may be tenuous or difficult, if not flatly refused (Piersel & Gutkin, 1983). The Classroom Check-Up (CCU; Reinke & Merrell, 2008) for example, is a class-wide consultation model used to address the need for classroom-level support while minimizing the treatment integrity problems common to SBC. Although the CCU represents an effective, evidence-based intervention targeting improvement in teacher classroom management skills, the training, time, and resources needed for its implementation may be impractical for some districts to implement, particularly on a large scale.

The challenges evident in the current professional development and consultation practices further illustrate a need for additional interventions, including consultation practices, targeting teacher classroom management behavior. Regular feedback around classroom management behavior in a quick, efficient, and ongoing format may offer a potential solution. Performance feedback interventions are predicated on the assessment of targeted observable behaviors, in this instance classroom management, followed by the delivery of these assessment results back to an individual. Currently, the few options available for easily and efficiently assessing classroom management make it difficult to provide performance feedback in this area. strengthening the connection between assessment and intervention, a renewed emphasis on hypothesis testing, on-going data

collection, and data-based decision making (Hintze, Volpe, & Shapiro, 2002) have complicated intervention around classroom management by requiring the application of a problem-solving approach to educator practices. This is to say, data-collection, hypothesis generation and testing, intervention, and data-based decision making focusing explicitly on educator behavior is a departure from the traditional student-focused interventions that have dominated scholarly work and applied practice historically. To supplement available classroom management assessments, The Direct Behavior Rating-Classroom Management (DBR-CM) was developed to give researchers and practitioners a technically adequate, efficient, and useful assessment of teacher classroom management behavior. The DBR-CM was designed to efficiently assess teacher classroom management behavior to facilitate attention, research, intervention development, and data-based decision making around classroom management.

Like Direct Behavior Rating (DBR; Christ, Riley-Tillman, & Chafouleas, 2009) for student behavior, the DBR-CM allows users to identify and quantify classroom management, as well as guide intervention efforts and monitor progress in response to intervention efforts. DBR assesses behavior using brief ratings of target student behavior following a specified observation period (Chafouleas, Riley-Tillman, & McDougal, 2002). Although DBR appears similar to individual items on a behavior rating scale, DBR has three unique and specific characteristics. DBR ratings occur in close temporal proximity to the target behavior, ratings are typically completed by a person who has direct experience with the target behavior of measurement during the period of interest, and ratings require a minimal level of inference (Christ et al., 2009). DBR combines the immediacy of systematic direct observation, the efficiency of general outcome measures,

and the psychometric reliability of behavior rating scales to repeatedly, defensibly, flexibly, and efficiently assess behavior (Christ et al., 2009). This unique combination of behavior assessment methodology builds on the strengths of systematic direct observation (SDO; e.g., repeatability, sensitivity to change) and behavior rating scales (e.g., psychometrically sound) to assess behavior in an efficient and direct manner, which makes it a highly desirable assessment option for both researchers and practitioners (Christ et al., 2009). Use of DBR methodology to develop an additional assessment tool for classroom management is an advantageous application. It combines the strengths, while limiting the weaknesses, of direct observation and rating scales (frequently used methods of behavior assessment) into a more efficient and usable form (Riley-Tillman et al., 2005).

DBR utilizes the general outcomes measures (GOM) approach to assess behavior in a manner similar to that used by Curriculum Based Measurement (CBM; Deno, 1995). GOMs combine a variety of specific behaviors that share traits into a single measurable domain (Deno, 2003; Shinn & Shinn, 2002). Consequently, each domain represents a group of behaviors that have been found to correlate with student success in the educational setting. Much like an oral reading fluency score encompasses several individual reading components, academic engagement, for example, can include a variety of specific behaviors required for or associated with focusing on instructional activity. Academically engaged to an outside observer could look like focusing on a teacher's instruction, reading silently, completing a worksheet, or discussing a topic with a peer. Each of these specific behaviors could be observed and assessed individually, but for efficiency in assessment, they can be easily grouped together as academic engagement. In

recent years, researchers have applied the same efficiency, practicality, and defensibility present in data-based decisions using CBM (Fuchs, Fuchs, & Maxwell, 1988) to observable student behavior. For instance, Direct Behavior Rating Single Item Scales (DBR-SIS) now serve as a well-established behavior assessment tool using brief ratings of target student behavior following a specified observation period (Chafouleas, Riley-Tillman, & McDougal, 2002). Research has indicated high levels of agreement in the brief estimates of these behavioral categories using Direct Behavior Rating, Single Item Scales (DBR-SIS) and the systematic direct observation of individual behaviors for students (Christ et al., 2009). This agreement suggests use of these general outcome measures for both academic (e.g., CBM) and behavioral assessments (e.g., DBR-SIS) are valid measures on which to base a variety of data-based decisions, while representing a significant savings of time and resources (Chafouleas, Riley-Tillman, & Sugai, 2007; Riley-Tillman, Chafouleas, & Briesch, 2007; Riley-Tillman, Christ, Chafouleas, & Boice-Mallach, 2011; Riley-Tillman, Kalberer, & Chafouleas, 2005).

The DBR-CM allows educators and scholars to efficiently, flexibly, and reliably assess teacher classroom management behaviors. Development of the DBR-CM addresses the limitation of assessment tools available for use within a problem-solving oriented approach to addressing educator performance in schools. In particular, the DBR-CM may contribute to consultative relationships seeking to improve teacher classroom management behavior. A frequent component of such relationships is performance feedback. Unfortunately, typical consultation models focusing on classroom management such as the CCU, may be too time and resource intensive for schools to employ. Fortunately, research suggests that a less coaching oriented approach to the provision of

performance feedback may also affect positive behavioral change. Fuchs and colleagues (1991) found that regular performance feedback in the form of graphic representation of curriculum-based measurement scores, resulted in significant increases in student achievement as well as changes in teacher instructional practice. These researchers found that intervention (e.g. instructional change as guided by formal consultation) and exposure to assessment data in the absence of an intervention (e.g. simple performance feedback), resulted in significant improvement in performance relative to a control group (e.g. no intervention or performance feedback). It should be noted that subjects in the performance feedback with a formal consultative relationship group outperformed the performance feedback only group. These results align well with the tiered service delivery and intervention models now adopted by many schools. From a tiered service delivery approach, the performance feedback only intervention represents a Tier II intervention. Should subjects not respond as desired to the performance feedback intervention, the consultation component could be added making the intervention more intensive. The application of this more time and resource intensive intervention following non-response to a Tier II intervention is consistent with Tier III intervention practices.

The study conducted by Fuchs and colleagues around performance feedback and consultation serves as a model for classroom management intervention using these intervention strategies. Given the success with student performance on CBM, if applied to teacher classroom management behavior as the targeted outcome, similar findings could be expected. When deficits in classroom management behavior are evident, a performance feedback intervention without a formal or significant consultation could be employed as a first, Tier II, intervention. In the event this intervention was determined to

be ineffective in producing desired behavior change, a more intensive intervention could be initiated. Like the Fuchs' study (1991), this intervention would employ performance feedback as part of a formal consultation relationship similar to the CCU. Assessment data DBR-CM would facilitate performance feedback in both interventions. This study was conducted to explore the effectiveness of a performance feedback without consultation or coaching intervention on teachers that appear to have Tier II level classroom management performance deficits.

The goal of the present study is to evaluate the effectiveness of performance feedback without formal consultation or coaching with data collected using the DBR-CM on classroom management behaviors. Specific research questions were (a) does performance feedback, using the DBR-CM, increase the classroom management skills of participating teachers as indicated by increases in overall rate of praise use of participating teachers and (b) is the DBR-CM perceived as usable within a performance feedback model, in terms of feasibility and acceptability?

Chapter 2

Role of Teacher in Learning

In the educational setting, teacher behavior before (and response after) a student behavior is influential in the reoccurrence or absence of that behavior in the future. Students engage in behavior based on the function the behavior, which is most commonly to gain something from a teacher or avoid a demand placed on them. At times this avoidance or escape of a task is due to a skill deficit. Generally, likely due to limited training in the principles of functional behavior assessment and behavior modification (Oliver & Reschly, 2010), educational staff lack an understanding of their importance in influencing the antecedents and consequences of student behavior. This is the understanding that in order to efficiently and effectively change a child's behavior, modifications to the environment, including teacher behavior before and after a behavior occurs, are necessary (Bruer, 1993). For those that do not subscribe to the behaviorist orientation of learning and behavior development, social learning theory and cognitive learning theory also place a great deal of emphasis on the role a teacher plays in a child's learning and development.

In social learning theory, learning, which includes changes in behavior, is facilitated through concepts such as modeling and observational learning (Sims Jr. & Manz, 1982). This theory, based on learning within a social context, supports the importance of the teacher as a primary model in the learning and behavioral development of children. This approach assumes people, especially children, learn from the environment and seek acceptance from society by learning through influential models

(Sims Jr. & Manz, 1982). Emotional affinity, acceptance, and other social behaviors serve to influence the repetition or suppression of actions.

Albert Bandura emphasized cognition and information-processing capabilities (cognition) that facilitate social behavior. Bandura's extension of the social learning perspective emphasizes models as important sources for learning new behaviors and for achieving behavioral change in institutionalized settings (Sims Jr. & Manz, 1982). He theorized observational learning occurs in relation to three "models:" (a) a live model, (b) verbal instruction, and (c) a symbolic model. Within this perspective, teachers are often directly responsible for the first two, and frequently influential in the third. Teachers frequently serve as a live model or the actual person demonstrating a desired behavior (e.g., "use of kind words"). The majority of a teacher's daily activities involve verbal instruction. In schools that use a Positive Behavior Interventions and Supports system (PBIS; Sugai & Horner, 2002), teachers repeatedly describe a desired behavior in detail, instruct students in how to engage in the behavior, model this behavior for students, and allow students to practice the behavior with feedback. Another means by which teachers provide behavior models for students is through the use of varied media in classrooms. As the use of technology and media increase in society in general, teacher use of media in classrooms expands. The symbolic models to which children are exposed, are highly influenced by teachers, again illustrating the importance of teacher behavior in successful learning for students.

In these theories of learning, teachers also play an important role in terms of reciprocal determinism. This idea states that an individual's behavior is influenced by the environment and characteristics of the person and vice versa (Williams, 2005). This

aspect of social learning states the modeling process requires several essential steps including attention, retention, reproduction, and motivation. First, in order for an individual to learn something, he/she must attend (attention) to the features of the modeled behavior. Second, individuals need to be able to remember (retention) details of the behavior in order to learn and later reproduce the behavior. Next, in reproducing a behavior, an individual must organize his or her responses in accordance with the modeled behavior (reproduction). This ability can improve with practice. Finally, there must be an incentive or motivation driving the reproduction of the desired behavior. Even if all of the above factors are present, the person will not engage in the behavior without motivation (Sims Jr. & Manz, 1982). In the school setting, an environment in which children spend a significant portion of their early formative years, the teacher is a primary model responsible for exposure to other models. This further illustrates the need to focus explicitly on the critical role teacher behavior plays in influencing student behavior. It is often easier to focus on the behavior of one disruptive student or one challenging class, and is more difficult to become introspective and examine what a teacher can do to change interactions with the situation (Reese, 2007).

Classroom Management Importance

The link between teacher and student behavior is intuitive and important (Myers et al., 2011) and teacher behavior clearly has an impact on student behavior (Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008). The influence of teachers on children in terms of overall effect of long-term outcomes may be secondary only to the influence of a parent or caregiver (Hamre & Pianta, 2001; Pianta & Stuhlman, 2004). Ample evidence has established children in well-managed classrooms spend more time engaged in

academic tasks, progress at a more rapid pace, and demonstrate higher levels of academic achievement (Brophy, 1985). Classroom management skills, or teacher behavior essential for successful classrooms, have been shown to predict student engagement and social competence, as well as reduce disruptive behaviors (Marzano & Pickering, 2003). Well-managed classrooms are distinguished by teachers' ability to monitor student attention and performance, establish behavioral expectations, and consistently implement rules and procedures that prevent problems from occurring (Reinke et al., 2008, 2008; Simonsen, Fairbanks, Briesch, & Sugai, 2006). La Paro & Pianta (2003) note well managed, optimally performing classrooms are characterized by low levels of conflict and disruptive behavior, smooth transitions from one type of activity to another, appropriate expressions of emotion, respectful communication and problem solving, strong interest and focus on task, and supportiveness and responsiveness to individual differences and students' needs. Implementation of evidence-based classroom and behavior management practices has been consistently shown to be associated with better student outcomes (Simonsen, Myers, & DeLuca, 2010). Unfortunately, teachers appear to rely more on unsystematic management strategies based on their own experiences rather than evidence-based techniques (Brophy & McCaslin, 1992). Consequently, when a teacher lacks good classroom management behavior, or the skill or resources to effectively manage the challenges within the particular context of their school and classroom, lower levels of on-task behavior and performance are observed (Marzano & Pickering, 2003).

Training in Classroom Management

In spite of evidence that effective classroom management leads to positive academic and behavioral outcomes for students, teachers receive surprisingly little

training in classroom and behavior management (Oliver & Reschly, 2010). Pindiprolu and colleagues (2007) found training programs spend relatively little time instructing pre-service teachers in the principles of effective classroom management, and only a small percentage of teachers feel adequately trained in this area. This may be due to general education teachers reports of having received little to no training in working with students with emotional and behavioral difficulties (Wagner et al., 2006). Similarly, Oliver and Reschly (2010) found only 27% of university special education programs taught classroom management as an independent course. This study also found when behavior management strategies were taught within other courses, the majority of the instruction involved responding to problem behaviors displayed by individual students (Oliver & Reschly, 2010). Both general and special education teachers reported that their biggest training deficit area was in preventing or intervening in problem behavior (Pindiprolu et al., 2007). Given the current policy climate related to teacher performance, and the need to employ evidence-based practices to improve social and academic difficulties, it is critical that teachers utilize classroom management strategies that effectively prevent problem behavior, improve pro-social behavior, and increase students' engagement with learning tasks.

Given the apparent opportunities for improvement in pre- and in-service training, it is not surprising school staff show little awareness of the role environment, or the adults around a student and their behavior, play in the development of problematic behavior and in turn affecting behavioral change (Sugai et al., 2000). Teacher behavior is fundamental in developing, shaping, and affecting change in student behavior, yet teacher

behavior is in large part ignored in school prevention and intervention planning (Sugai et al., 2000; Tillery, Varjas, Meyers, & Collins, 2010).

Professional Development Activities and Classroom Management

Upon completion of their formal university training and student teaching, further training that results in durable changes in teacher classroom and behavior management is sparse (Corcoran, 1995; Elmore, 2005). This may be the result of the challenge faced by administrators to identify effective, relevant, and feasible ways to improve teacher behavior when current levels of performance are not optimal (Myers et al., 2011). This is particularly problematic in instances where training programs neglect behavior and classroom management as points of curricular emphasis, a common complaint of teachers (Pindiprolu et al., 2007). The scope and volume of research dedicated to developing, validating, and assessing interventions targeting students and student outcomes far exceed those devoted to teachers and teacher outcomes. Typically, school building or districts use the “in-service approach” or mentorship model to provide staff with professional development opportunities (Corcoran, 1995; Guskey, 2000). In-service training is typically delivered in relatively passive, one-session workshops with few opportunities to achieve skill fluency through practice or feedback. Unfortunately, this approach is often ineffective, and typically more informational than developmental (Corcoran, 1995; Elmore, 2002). Frequently, school districts assign new teachers to a mentor, but the purpose, goals, and further professional development within this mentorship vary greatly from district to district and individual to individual (Feiman-Nemser, 1996; Huling & others, 2001). School districts and buildings often devote several days to professional development and in many cases have weekly professional

learning community (PLC) meetings (DuFour & Eaker, 1998; Stoll, Bolam, McMahon, Wallace, & Thomas, 2006). In each of these instances, it appears these efforts rarely result in long-lasting change in teacher practices (Elmore, 2002). To address shortcomings in these approaches to teacher skill development, efforts are shifting from one-time, in-service style training events to more systematic and continuous teacher support practices (Sugai et al., 2000).

An additional approach to improving student outcomes is seen in an emphasis on treatment integrity, as noted in literature focusing on treatment integrity or treatment fidelity within a multi-tiered system of support (MTSS) or response to intervention (RTI) model. Broadly, treatment integrity represents the degree to which an intervention (treatment) is implemented as designed (Gresham, 2008; Gresham, 2008; Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000). Evidence-based interventions often have specified procedures. More specialized definitions of treatment integrity use a multidimensional framework consisting of five components within treatment integrity: (a) adherence, (b) quality of delivery, (c) program differentiation, (d) exposure, and (e) participant responsiveness (Dane & Schneider, 1998; Hagermoser Sanetti & Fallon, 2011). Adherence, the dimension of treatment integrity that may be most easily assessed, is measured through self-report and, less frequently, direct observation. Percentages of adherence are calculated using the procedures stipulated by the intervention protocol and the procedures actually executed. Proponents of a treatment integrity, also commonly labeled implementation fidelity, approach, argue that by virtue of implementing interventions for students, teacher behavior is also modified. This modification in teacher behavior to address a student's needs theoretically exposes a teacher to practices that

could then be generalized to other students or contexts. While this could result in long-term changes in teacher behavior, research has yet to fully explore the effects of treatment integrity on the enduring adoption of intervention practices (generalization) by teachers (Corcoran, 1995; Gresham, 2008; Gresham et al., 2000). A survey conducted by Cochrane and Laux (2008) indicted the overwhelming perception that assessing treatment integrity, as part of the intervention implementation process, is essential. However, this same study found that treatment integrity is rarely assessed in individual consultative and team-based intervention efforts (Cochrane & Laux, 2008). A treatment integrity approach perpetuates the belief that the problem is within the student rather than the environment around the student. While subtle, this may be extremely important in perpetuating systemic beliefs, intervention goals, and the focus of school-based interventions research. It is believed that an approach to intervention that explicitly targets teacher behavior as outcomes may result in better outcomes for more students (Reinke et al., 2008).

Treatment Integrity and Classroom Management

In applied school settings, treatment integrity is frequently evaluated in two main ways: (a) by determining various adherence percentages to the main aspects of an intervention, and (b) through student outcome measures (i.e., if the student outcome improved, treatment integrity must have been good; (Lane, Bocian, MacMillan, & Gresham, 2004). Within scholarly research, evaluation of treatment integrity may become more complex (Hagermoser Sanetti & Fallon, 2011), but a more nuanced approach to its evaluation appear frequently neglected in day-to-day school-based intervention efforts (Cochrane & Laux, 2008; Lane et al., 2004) Intervention components typically measured

include number and frequency of implementation sessions, and the steps or core components of an intervention adhered to versus the number specified by the intervention protocol (Gresham, 1989; Gresham et al., 2000). Within these dimensions of treatment integrity assessment, emphasis is placed on frequency aspects of intervention implementation while neglecting quality. In addition to continuing to place the emphasis for a need for change on the student, this approach also deemphasizes skill acquisition and generalization in favor of systematic, robotic delivery of intervention steps as stipulated by intervention protocols. It could be argued treatment integrity in general, pays little attention to ensuring intervention implementers, frequently teachers, have or develop the skills necessary to implement interventions with any specified degree of quality. This has prompted researchers to explore this aspect of treatment integrity specifically (Hagermoser Sanetti & Fallon, 2011), as it appears to be a shortcoming of a treatment integrity approach when viewed as a means of improving teacher skill, particularly as it relates to classroom management. A goal of most skill development is generalization, or the application of a learned skill to a new environment or situation. At the very least it emphasizes temporary behavioral change in teachers, requiring adherence, or variation in behavior, to intervention steps until the intervention is completed or until the student difficulty is remediated. Within a treatment integrity approach, when teacher adherence to protocol steps is poor, if assessed at all, there is often little external accountability (Hagermoser Sanetti & Fallon, 2011). Within this approach, responsibility for implementing protocol steps falls on the teacher and systematic assessment and, if needed, remediation of teacher implementation adherence is frequently neglected.

Performance Feedback

Given initial training by itself does not appear to adequately prepare teachers (McKinney et al., 1983) and subsequent professional development may be ineffective in addressing teacher training needs (Elmore, 2002), schools should consider systematic, on-going, or formative performance feedback to promote desired teacher classroom management behavior. On-going, data-based feedback on current performance of targeted behaviors is among the most promising methods for increasing the implementation of evidence-based instructional practices within a school based consultation model (Reinke, Lewis-Palmer, & Martin, 2007). Performance feedback is defined in the school consultation literature as “monitoring a behavior that is the focus of concern and providing feedback to the individual regarding that behavior” (Noell et al., 2005). Feedback, as a more general construct, has been studied extensively in education (Hattie & Timperley, 2007; Timperley, 2008) and several studies have documented the effectiveness of performance feedback on behavior change with teachers (D. Fuchs & Fuchs, 1993). A closer look at this research reveals a fuller picture of performance feedback within. Currently, common applications of performance feedback for teachers can be seen in evaluations of treatment integrity for a specific intervention with a child or small group, formative evaluations conducted by building or district administrators, or as part of the consultation process. In each application the performance feedback provided combines quantitative feedback (incorporating numeric information collected using an assessment tool) with qualitative feedback or coaching designed to elicit desire to change or develop skills within a teacher (see Reinke et al., 2008; Reinke, Stormont, Webster-Stratton, Newcomer, & Herman, 2012; Sprick, 2013). Though promising alternatives in

some regards, shortcomings are noted in the current applications of performance feedback.

Difficulties with Performance Feedback

Consultation-, summative evaluation-, coaching-, or treatment integrity based performance feedback approaches attempt to incorporate an accountability component in intervention implementation and have become popular in the school-based literature as methods of increasing teacher performance (Solomon et al., 2012). When performance feedback is employed to increase the fidelity with which a prescribed intervention is implemented, across a continuum of tiered interventions for both behavioral and academic problems, the additional layer of accountability would theoretically improve teacher practice. As discussed previously, this scenario emphasizes the problem within student perspective. While beneficial in individual instances, the intervention for an individual student that results from this approach deemphasizes teacher skill development and generalization to all students. This approach continues to promote the shortsighted approach of intervention focusing on individual student skill deficits, rather than improving the environment and the individuals making up the environment in which all children are learning.

Summative, evaluation-based performance feedback for teachers by administrators is also rife with shortcomings (Peterson, 2004). First, teachers are often aware far in advance of these evaluations and can prepare accordingly by adjusting performance to meet the expectations of observers. This would allow teachers to prepare lessons, coach students, and exhibit desired behavior in a manner that could be discrepant from their typical performance (Peterson, 2004). Additionally, these evaluations occur

infrequently, providing little useful information regarding small increments of change over time. Often teacher evaluations focus on minimum acceptable levels of performance. The instruments used to collect data and provide feedback typically lack normative information or standardization. These evaluations are predominantly subjective in terms of the content or areas targeted for evaluation, as well as the individual conducting the evaluation. Also, the feedback may or may not be useful in terms of making meaningful changes to classroom practices. This performance feedback is often linked to ratings used to make decisions about contract retention based on minimum levels of acceptable performance. Taken together, this model of performance feedback yields little information to make meaningful changes to day-to-day practices of a teacher in the classroom (Murnane & Cohen, 1986).

Advances in Performance Feedback

In recent years, researchers have proposed several SBC models. The Functional Assessment of Academic Behavior was developed as a tool for consultants to assess the instructional setting and needs of individual students with learning problems (Ysseldyke and Christenson, 2002). This model includes teacher expectations, classroom environment, and parent participation components to foster student achievement and adaptation. Erchul and Martens (2002) proposed a school consultation model for school psychologists that integrates elements of mental health and behavior consultation, as well as literature from their own relational communication research. These modern models of instructional consultation share an emphasis on careful assessment of instructional domains to guide intervention development, and they situate the learner as part of the instructional system rather than as the primary source of the problem.

The Classroom Check Up (CCU)

A persistent limitation within existing models is the continued focus on students as targets of interventions (e.g., intervention X will increase student engagement) rather than on changing the entire classroom ecology (Sheridan, Welch, & Orme, 1996). Rather than attributing student difficulty to lack of exposure, modeling, practice, or instruction, student focused interventions promote that difficulty is due to core ability deficit within a child. The “problem within-student” perspective perpetuates the belief that academic, social, or behavioral problems are the result of disordered characteristics within a student rather than the result of deficient environmental experiences. This approach, similar to a traditional approach to identification and placement in special education, can be stigmatizing (Florian, 2007). A focus on individual difficulties rather than the environmental factors that, at the very least confound performance difficulties, may be inefficient and stigmatizing. The cost-benefit ratio is likely better for interventions that will ultimately benefit more students and prevent future problems from occurring. This efficiency could also be evident in the time and money invested in intervention implementation for students whose needs are not met by these universal, preventative practices. The revisions to the Individuals with Disabilities Education Act IDEA in 2004 and the adoption of RtI approach to service delivery have attempted to shift the focus of the education providers away from the problem within-student perspective (D. Fuchs & Fuchs, 2006; Yell, Shriner, & Katsiyannis, 2006). Intervention efforts focusing on environmental factors (e.g., curriculum, classroom management, school-wide behavior supports) have the potential to address the needs of all students rather than an individual child (Reinke et al., 2007, 2008; Sugai & Horner, 2002).

The CCU is a consultation model intended to address the need for classroom-level support by focusing on teacher skill development in a SBC approach. The CCU builds on existing consultation models in that it emphasizes class-wide change and motivational enhancement strategies that are informed by an extensive social-psychological literature (Reinke et al., 2008). The CCU uses a five-step process in which a consultant assess the classroom, provides feedback, provides a menu of intervention options, supports intervention selection, and guides self-monitoring. Through these steps the essential components of the CCU are addressed. These components include: Teacher interview, Classroom ecology checklist, Classroom observations, Consultant provides feedback on assessment findings, Feedback including strengths and weaknesses, Teacher and consultant collaboratively develop a menu of options to intervene to create positive classroom outcomes, Teacher chooses interventions to implement, Consultant provides ongoing support in the implementation of the intervention(s), and Teacher monitors daily implementation of the chosen intervention using an intervention procedural checklist (Reinke et al., 2011). The CCU incorporates some specific motivational enhancement strategies, including rapport building, motivational interviewing, personalized feedback to teachers on classroom behaviors, encouraging personal responsibility for decision making while offering direct advice if solicited, development of a menu of options for interventions, and supporting teacher self-efficacy by identifying existing strengths and times when teachers have successfully changed classroom behaviors in the past (Reinke et al., 2011; Reinke et al., 2008).

Although the CCU represents an effective approach to an intervention designed to develop teacher skill improvement or acquisition, the time and resources needed for

implementation may be impractical on a large scale. The CCU may not be as accessible to most districts, buildings, or teachers. The current economic climate in general, and within education specifically, necessitates frugal expenditure of district funds. This may mean that funding for consultants is limited. The time required to implement the CCU may be beyond that which teachers and administrators have or wish to devote to this type of ongoing professional development. Each of the steps in the CCU has the potential to be very time consuming and may be impractical for some school staff. Teachers are typically hard-pressed to find time for extracurricular activities. To a lesser degree, the amount of time needed to conduct the assessments and data collection needed for the CCU may be impractical for schools. The CCU incorporates teacher monitoring of intervention implementation, but also requires ongoing feedback from the consultant. This again illustrates a potential strain on time and resources that may not be manageable in the school setting, but at the very least pose an additional burden for teachers or consultants. Another factor to consider is teacher resistance to a consultative relationship. Much of the literature around school-based consultation characterizes this process as collaborative. Collaboration, denotes openness, willingness, cooperation, and equality of power within the consultant-consultee (teachers) relationship. While this may be true of consultation targeting an individual student or small group of students, this portrayal may not be accurate for consultative relationships with teacher behavior change as the targeted outcome. These interactions are frequently something that is stipulated by a building or district administrator rather than sought out by a consultee (Piersel & Gutkin, 1983; Truscott et al., 2012). For staff such as School Psychologists, Behavior Specialists, PBS/PBIS coaches, peer mentors, and administrators, engaging in a consultative

relationship with teachers can be tenuous if not flatly refused. School based consultants find themselves faced with the difficulty of working with someone that is content with their own functioning and genuinely believes the problem lies within a student or group of students. Issues around teacher resistance to consultation and behavior change are evident in the development of the CCU. Developers of the CCU infused components of motivational interviewing to overcome consultee resistance to change (Reinke et al., 2011). Such an approach has been effective in facilitating change, however, as noted previously, it may be too time and resource intensive in some contexts. The challenges to the current professional development and consultation practices illustrate a need for a more practical, easily accessible intervention for classroom management. A basic performance feedback loop for classroom management may offer such a solution. A method of providing performance feedback to teachers around classroom management that is independent of the coaching component often associated with performance feedback interventions may be beneficial, as performance feedback has shown promise in improving teacher performance in this area.

Assessment for Performance Feedback

Performance feedback is predicated on the ongoing collection and use of assessment data of a target behavior. Currently, the number of assessments dedicated to teacher classroom management behavior is limited. Available information suggests classroom management is often assessed through student outcome data. A review of PBIS (Horner, 2000) recommendations for assessment indicate few additional options. PBS strategies require use of practices consistent with positive classroom management practices, such as clear communication of behavioral expectations and use of praise.

Available information suggests PBS proponents rely heavily on educator self-report and student outcome data to determine adherence and effectiveness of school- and classroom-wide practices. Recommended evaluation data includes office discipline referrals (ODR) or other school wide information system data (SWIS; May et al., 2000). Additional data sources include teacher requests for assistance, teacher ratings and rankings of students, teacher referrals to special education, informal “walk-throughs,” the Classroom Management: Self-assessment Revised (Simonsen et al., 2006), and direct observation of frequency, duration, and severity (May et al., 2000; Simonsen et al., 2006). The emphasis on student outcomes rather than teacher outcomes is also reflected in other examinations of school- or classroom-wide interventions. In their evaluation of PBIS implementation in Maryland, Barrett and colleagues (2008) identified their two outcome variables as ODR and suspension rates. Further, a recent meta-analysis of classroom management techniques found that all 26 studies included used student performance as the outcome variable (King, under review). This review of classroom management literature suggests that teacher behavior is often viewed as treatment integrity data rather than a true outcome variable. Assessment tools targeting teacher classroom management behavior appear to be limited to direct observation of frequency, rate, and duration information or self-report information of adherence to treatment protocols.

Systematic Direct Observation

Within educational settings, significant effort in the last two decades has been dedicated to shifting from a diagnosis-driven model to a problem solving model (S. L. Deno, 1995). The shift to a data-driven, hypothesis testing, and decision making approach within educational systems has increased the need to link assessment and intervention

(Hintze et al., 2002). To guide the problem-solving process, educators need assessment tools and procedures that are technically adequate, efficient, and useful (Fuchs & Fuchs, 2006; Riley-Tillman, Chafouleas, Briesch, & Eckert, 2008). In research and practice, when behavior is the variable in question, systematic direct observation (SDO) is often the means by which performance data is collected. Within research, SDO is considered the gold standard of behavior assessment measures (Riley-Tillman et al., 2011), as it provides reliable and accurate measurements of target behavior (Cone, 1978). SDO of behavior has also become a prevalent method used to collect data by practitioners as well (Reschly, 2003). With SDO, an observer records the frequency, duration, or severity with which an operationally defined behavior occurs (Salvia, Ysseldyke, & Bolt, 2007).

SDO typically refers to a set of techniques that observers use to quantify behavior using one or a combination of dimensions (Kratochwill, Alper, & Cancelli, 1980). Because behavior can be quantified using various dimensions, a clear operational definition is essential to determining which dimension and subsequent observation technique is appropriate. Operational definitions must be objective to ensure that specific instance of behavior can be measured as well as to allow others to replicate this measurement (Hintze et al., 2002). Operational definitions should identify the target behavior explicitly, identify what the target behavior is not, and distinguish the target behavior from similar behaviors (Hintze et al., 2002). Once an operational definition is achieved, an SDO observation procedure can be selected.

Frequency or event recording data is a common SDO assessment. This collection procedure involves counting the number of discrete occurrences of a targeted behavior in a specified observation period. An example of frequency data collection could be tallying

the number of times a student talks out during a math lesson. A second SDO data collection method is duration recording. Duration recordings measure how long a targeted behavior lasts. An example of duration recording data collection is tracking the number of seconds between the beginning and end of child crying. Latency, another SDO data collection option, measures the time that lapses between a stimulus or direction and the occurrence of the desired behavior. An example of latency data collection may be counting the number of seconds between presenting a child a verbal direction to open her text book to page 87 and her opening her book to the desired page.

To address some difficulties inherent in collecting frequency, duration, and latency data in the applied setting, SDO may incorporate interval recording procedures. Time sampling interval recording involves selecting a time period for observation, dividing the period into equal intervals, and recording the presence of the targeted behavior in each of the intervals (Merrell, 2001). Whereas frequency, duration, and latency observations provide exact rate or duration information, interval recording procedures provide approximates or percentages of behavior occurrence (Hintze et al., 2002). Time-sampling options include whole-interval recording, partial-interval recording, and momentary time-sampling. Whole-interval recording requires the target behavior to be present throughout the entire interval in order to be scored, whereas partial-interval recording scores the target behavior as occurring if it occurs in any part of the interval. Momentary time-sampling scores a target behavior as present or absent at the instant when the an interval begins (Hintze et al., 2002).

Despite the utility of SDO for accurate and reliable behavior assessment, several disadvantages exist when using SDO that significantly affect the feasibility of use in

schools or with larger-scale research projects (Riley-Tillman et al., 2007, 2005). A potential drawback to the use of SDO stems from the taxing nature associated with use in an applied setting. SDO is often a time-intensive method of data collection, requiring focused attention of an observer for each observation period. Barring the use of potentially intrusive and costly video recording technology, an observer must be present in the environment in which the behavior occurs to conduct SDO and only SDO. Often, multiple observations are required when assessing behavior, thus requiring a significant amount of time to assess one subject (Riley-Tillman et al., 2007, 2005). SDO may also require significant training which makes it time intensive for researchers and impractical for practitioners (Riley-Tillman et al., 2011). The presence of an observer (or video recording equipment) may result in atypical behavior from individuals in that environment (Riley-Tillman et al., 2011). In these instances, though SDO may provide accurate measurement of the target behavior, because of the influence of the observer in the environment, the data collected may not be representative of true functioning. The use of assessments that limit these Hawthorne Effects are advantageous for both researchers and practitioners (Thompson, 2006). The advantages of SDO make it an excellent assessment option in many situations. However, users must weigh the disadvantages of using SDO with the advantages as they select assessment tools. In any case, it is advantageous to have a variety of data collection options that to provide alternatives for SDO in situations when it may not be a feasible measurement method.

Classroom Management Assessment

As of late, some researchers have identified teacher classroom management as a an outcome variable for intervention efforts (see Reinke et al., 2014; Reinke, Lewis-

Palmer, & Merrell, 2008; Reinke, Herman, & Stormont, 2013; Reinke, Herman, Stormont, Newcomer, & David, 2013). Many of these studies utilize the Multi Option Observation System for Experimental Studies (MOOSES; Tapp & Wehby, 1993; Tapp, Wehby, & Ellis, 1995), a time sequenced SDO program, to record a variety of teacher behaviors as outcome variables. This observation system represents an improvement to paper and pencil direct observation recording procedures, but may lack the capability to measure dimensions of classroom management that cannot be captured by frequency or duration. As with SDO generally, this direct observation system has the potential to be cumbersome for use by educators in schools. The MOOSES system is costly (\$1000 for a site license) and requires extensive training to step up observation variables, ensure reliability in use, and to interpret data collected. This observation program can also become complex as it allows for several variables to be coded within an observation.

Difficulties with contextual appropriateness and usability are also evident in another frequently used assessment option for classroom management, the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008). The CLASS expands beyond frequency and duration data to rate other dimensions of teacher-student interactions believed to be essential for student learning and development. This observation system has a heavy emphasis on behaviors consistent with common operational definitions of classroom management. This assessment focuses on the breadth, depth, and frequency of teacher-student interactions to provide data that can be used to guide improvement of classroom management skills (Pianta et al., 2008). Though the CLASS system provides useful information that traditional direct observation may not, CLASS observers must spend extensive time attending costly trainings to learn

definitions of domains observed and scoring procedures for the instrument. Thus, it may not be practical for some school districts, buildings, and practitioners.

Alternative Classroom Management Assessment

To address the limitation of current classroom management assessment options, an assessment tool was created incorporating strategies and approaches embraced by general outcomes measures. Development of this assessment sought to efficiently assess essential teacher classroom behaviors, or classroom management, in an efficient, effective, and practical manner. There is a wealth of research supporting the use of CBM as a progress monitoring assessment practice to make data based decisions in the educational setting. For example R-CBM (Reading-Curriculum Based Measurement, a measure of oral reading fluency or words read correctly in one minute) has substantial empirical support for its validity as an overall indicator of general reading competence (Ardoin, Christ, Morena, Cormier, & Klingbeil, 2013; Christ, Zopluoglu, Long, & Monaghan, 2012; Christ, Zopluoglu, Monaghan, & Van Norman, 2013). Increasingly, R-CBM has been of interest to general education teachers and school administrators, in part, because of its prevention focus (Shinn, Shinn, Hamilton, & Clarke, 2002) and its consistently strong relationship to statewide, high-stakes reading tests (McGlinchey & Hixson, 2004). Given its unique combination of efficiency, low cost, and validity with respect to important educational outcomes, R-CBM is a worthwhile indicator for judging the effectiveness of overall reading instruction and intervention support. In recent years, researchers expanded the efficiency, practicality, and defensibility of data-based decision using CBM to behavior by developing Direct Behavior Ratings Single Item Scales (DBR-SIS).

Direct Behavior Rating Background

DBR assesses behavior using brief ratings of target student behavior following a specified observation period (Chafouleas, Riley- Tillman, & McDougal, 2002). DBR is not the first assessment tool of this kind. Tools such as the home-school note, daily behavior report card, daily progress report, and good behavior notes share similar methodology and features (Chafouleas, Riley-Tillman, & McDougal, 2002). Though traditionally this assessment methodology had been applied to intervention efforts, Chafouleas and colleagues (2007, 2009, Christ et al, 2009) examined the application of the methodology to assessment. DBR is unique from behavior rating scales in that assessment occurs in close temporal proximity to the target behavior, ratings are typically completed by a person who has direct experience with the target behavior of measurement during the period of interest, and ratings require a minimal level of inference. Additionally, DBR combines the immediacy of systematic direct observation, the efficiency of general outcome measures, and the psychometric reliability of behavior rating scales to repeatedly, defensibly, flexibly, and efficiently assess behavior (Christ et al., 2009). This suggests DBR may be the most appropriate option for assessing observable behaviors, as it yields data that is as reliable and valid as SDO and behavior rating scales, in a more efficient, direct manner. This unique combination forms a method of behavior assessment that builds on the strengths of SDO (e.g., repeatability, sensitivity to change), behavior rating scales (e.g., efficient data collection), and GOM (e.g., efficient generalizable data collection) to assess behavior (Chafouleas, 2011; Christ et al., 2009). Defensibility is the psychometric strength of data in relation to intended purpose (Chafouleas, 2011). Highly standardized and resource intensive assessment tools

exemplify defensibility. Though these tools may possess valuable psychometric strengths in the evaluation of behavior problems, the costs often outweigh the benefits of their use. For this reason, defensibility is not the only consideration for assessment development or selection, and usability must also be considered. Usability refers to considerations surrounding acceptability, understanding, feasibility, and systems support (Chafouleas, Briesch, Riley-Tillman, & McCoach, 2009). Together, defensibility and usability are critical pieces considered in the selection of an assessment instrument (Chafouleas, 2011).

General Outcome Measures Background

General outcome measures (GOMs) have been used for several decades by researchers to gather defensible and usable data within a problem solving approach to service delivery (Shinn, 2002). GOMs are useful for dual purposes involving establishing (screening) and evaluating progress (progress monitoring) toward key outcomes, both of which serve as foundations of problem-solving assessment. Administration of GOMs serves to identify students with performance deficits that can then be targeted for additional support. GOMs do not provide detailed diagnostic information, but rather a good indication of deficit or risk for failure. Often, after additional supports are implemented, the GOM is administered more frequently to monitor the progress of the intervention. The most prominent application of GOMs in education is seen in Reading Curriculum-Based Measurement (R-CBM; Shinn & Shinn, 2002). R-CBM uses the principles of GOM by combining a variety of specific reading behaviors into a single outcome measure. Although GOM principles have been applied liberally to academics, they have not been applied as readily to social behavior (Christ et al., 2009).

Direct Behavior Rating-Single Item Scales Background

DBR researchers used the principles of GOMs in combination with DBR methodology (temporal proximity, direct observer completes rating, and minimal inferencing required), to develop DBR Single Item Scales (DBR-SIS). Initial research yielded three constructs that encompassed many observable student behaviors associated with favorable student outcomes. These constructs are academically engaged, disruptive, and respectful. Like an R-CBM score encompasses several individual reading components, DBR-SIS combines a variety of specific behaviors into the single outcome constructs of academically engaged, disruptive, and respectful. Academic engaged for example, includes a variety of behaviors that individually could be operational definitions of academic engagement. Academically engaged to an outside observer could look like focusing on a teacher's instruction, reading silently, completing on a worksheet, or discussing a topic with a peer. Each of these specific behaviors could be observed and assessed individually, but for efficiency in assessment, they can be easily grouped together and viewed collectively as academic engagement. Research indicates high levels of agreement in the brief estimates of these behavioral categories and the systematic direct observation of individual behaviors for students, suggesting that the use of these general outcome measures for both academic and behavioral assessments are not only valid measures on which to base decisions, but also represent a significant savings of time and resources dedicated to assessment practices (Chafouleas, 2011; Chafouleas, Riley-Tillman, & Christ, 2009b).

Initially developed and evaluated through a systematic line of research, DBR-SIS assesses specific observable student behaviors in the classroom setting associated with

positive school outcomes by grouping them into domains (i.e., academically engaged, disruptive, and respectful; Chafouleas, Riley-Tillman, et al., 2009). Through this research, with over 30 peer-reviewed publications to date, DBR and DBR-SIS have been established as a defensible, flexible, efficient and repeatable methods of behavior assessment. DBR-SIS research has examined the influence of wording (i.e., positive vs. negative) and specificity (i.e., global vs. discrete definitions) of target behaviors on the accuracy of ratings (Riley-Tillman, Chafouleas, Christ, Briesch, & LeBel, 2009), the instrumentation of single-item DBR scales with regard to scale gradient (e.g., 0-5, 0-10; Christ et al., 2009), and the number of observations necessary for defensible use in low-stakes and high-stakes educational decisions (Christ et al., 2009). These findings indicate comprehensive wording of target behaviors and careful consideration specific example target behaviors increases accuracy in ratings. Findings resulted in recommendations for minimum scale gradients for adequate reliability (6 or more) and the minimum number of observations for low- (5 observations) and high-stakes (15 observations) decision-making. Taken together, these findings resulted in the development and refinement of DBR-SIS and provide guidelines for development of other assessment tools, procedures, and decision-making using DBR assessment methodology.

Direct Behavior Rating-Classroom Management Development

Following a review of the literature on effective classroom management, a GOM assessment was developed based on DBR methodology and DBR-SIS instrumentation. The Direct Behavior Rating-Classroom Management (DBR-CM) uses DBR assessment methodology to measure five broad behavior domains of classroom management. The resulting broad behavior domains were chosen based on prevailing scholarly literature

focusing on classroom management behaviors (see Brophy, 1983; Reese, 2007; Reinke, Lewis- Palmer, & Martin, 2007; Simonsen et al., 2008; and Snyder, 1998). Following the initial construction of the DBR-CM, broad behavior domains were refined by experts in both assessment and classroom management. Selected constructs included praise, communication, rapport, enthusiasm, and engagement. Each category includes numerous specific behaviors that are similar to one another and can be grouped together based on their commonality. Operational definitions for the broad behavior domains making up the DBR-CM are provided in Table 1.

The first broad construct of classroom management identified on the DBR-CM is Praise. Praise is operationally defined as the use of positive praise statements in response to the behavior and performance of students in the classroom and a visibly general positive attitude towards all students. Praise in the classroom looks like using more praise statements, ideally more behavior-specific praise than general praise, using praise contingent on expected behavior, providing three or more praise statements for every reprimand, displaying more positive than negative attitude when interacting with students, and having an overall tone that is positive and not negative or sarcastic. The purpose of praising students is to recognize or show interest in them, to encourage them, to describe what we observe in their behavior, and to evaluate their performance (Douglas Bartholomew, 1993).

The second classroom management construct assessed by DBR-CM is Communication. Communication is operationally defined as clearly conveying goals and expectations of a classroom and/or instructional period to students. Communication in the classroom looks like providing clear expectations for the students, clearly expressing

instructional objectives, offering opportunities for clarifying questions, clearly presenting behavioral expectations verbally and/or visually (this may include actively teaching behavioral expectations including classroom rules and routines), and a transition procedure that appears to be known and followed by majority of students. Good classroom management is based on students' understanding of their expected behaviors (Strout, 2005). It is therefore essential that teachers effectively communicate expectations.

The third construct assessed by the DBR-CM is Rapport. Rapport is operationally defined as is the quality of the student-teacher relationship, especially that of mutual trust, emotional affinity, acceptance, and positivity. In the classroom, rapport looks like a general feeling of mutual warmth and acceptance between the teacher and students. Rapport would be evident in the teacher's frequent use of children's names and positive interactions between the teacher and students. In classrooms where good rapport has been established, students appear to feel comfortable approaching the teacher or asking questions and the teacher appears to feel comfortable, positive, and genuine in his/her interactions with students. Teachers build positive classrooms by developing relationships with their students (Schaps et al., 1997).

The next broad category of classroom management identified on the DBR-CM is Engagement. Engagement is operationally defined as 90% or more of students being involved in, focused on, or participating in instruction and/or classroom activities, and posing questions to students occurs frequently. In the classroom, engagement looks like 90% of students are clearly academically engaged at all times, level of observable disruptions in the classroom is minimal, the teacher provides four or more opportunities

for students to respond per minute during instruction, and asking most of the students at least one question during instruction. Research has identified a strong positive relationship between academic learning time, defined as the portion of time students spend actively and productively engaged in learning, and student achievement (Gettinger & Seibert, 2002).

The final classroom management construct included on the DBR-CM targets Enthusiasm. Enthusiasm is operationally defined as the delivery of instructional content in a meaningful, memorable, and/or engaging manner. In the classroom enthusiasm looks like an upbeat tone and pace of the instruction being delivered; instructional content is being related to a familiar life or practical application, topic, or activity; and instruction incorporates varied activities (e.g., students as teachers, group work, pair and share, current event, etc.). When taken at face value, enthusiasm means simply doing something with excitement, fervor, or eagerness. Beyond the pace, energy, and excitement in a teacher's delivery of instruction, enthusiasm also encompasses the quality of the activities and examples used during instruction. Enthusiasm also means providing alternative examples that are relevant, practical, and relatable to a student's experiences, life, current events, or future. Students are most likely to perform in desirable ways when engaged in realistic, worthwhile tasks (Reed & McNergney, 2000).

The development of the DBR-CM should allow for more efficient collection of data targeting teacher classroom management skills, as classroom management is essential to promoting positive outcomes for students in the educational setting. DBR-CM data allows for the provision of performance feedback for teacher classroom management behavior in an efficient, flexible, and defensible manner. Thus far,

performance feedback within a SBC model represents an effective means of improving teacher behavior. Though performance feedback within SBC models has shown desired effects when employed in consultation models like the CCU (Reinke, 2008), such systems of consultation may not be feasible for some districts or buildings to implement. The time and resource intensive nature of such an approach may not be appropriate or necessary for some minor classroom management behavior performance deficits. Classroom management intervention efforts relying on performance feedback as a component of a SBC approach are consistent with Tier II or III intervention efforts. Performance feedback in the absence of a coaching or SBC relationship is more consistent with a Tier I intervention. The effectiveness of performance feedback on student academic performance in the absence of a formal intervention has been documented (see Fuchs et al., 1991), but such approaches have yet to be applied to teacher classroom management behavior. The DBR-CM aligns well to such an intervention approach as it, in keeping with DBR-SIS, is easily learned and used by raters. The less taxing nature of the data collection and dissemination of performance feedback with an intervention utilizing the DBR-CM to provide brief, graphic performance feedback without a formal SBC relationship to address Tier II level classroom management performance deficits worthy of scholarly attention. This study sought to evaluate the effectiveness of a performance feedback intervention via presentation of a line graph displaying DBR-CM scores. This classroom management focused intervention appeared to be less time and resource intensive than current performance feedback based interventions (e.g. coaching or consultation). Project coordinators hoped that performance feedback delivered in a simplified format could be

used to address mild to moderate classroom management performance deficits without requiring the time, money, and individuals typically needed for coaching or consultation oriented intervention efforts.

Research Questions

The overall goal of this study is to determine the effectiveness of performance feedback using the DBR-CM in improving classroom management behaviors. The development of the DBR-CM was spurred by a combination of influences, including inefficiency of current classroom management assessment tools, as well as the need for classroom management interventions. It was believed that the DBR-CM represents a means to provide teachers feedback on classroom management functioning in an easily accessible and meaningful manner. It is hypothesized that regular, simple performance feedback using the DBR-CM will result in an increase in total rate of overall praise observed in classrooms. The construction of the DBR-CM, consistent with GOM methodology, embeds many specific classroom management behaviors into parsimonious categories. It is hypothesized that brief exposure to these categories (and examples and non-examples of each category) and subsequent performance feedback, delivered via graphic representation of performance scores, will result in an increase in a measure of teacher classroom management behavior. Prior research has demonstrated the positive effect of performance feedback for specific classroom management behaviors (Reinke, Lewis-Palmer, & Martin, 2007, Reinke and Merrell, 2008). These specific behaviors are encompassed in the five categories adopted by the DBR-CM and thus, performance feedback using this instrument should result in similar behavior change. This study sought to make the performance feedback process more efficient by examining the utility

of an assessment tool developed using methodologies that yielded valid, defensible, and usable data when applied to other behaviors. Specific research questions this study seeks to answer are as follows:

1. Does performance feedback without a formal consultation relationship using the DBR-CM, increase participating teachers' overall rate of praise, a variable often used as an estimate of overall classroom management? It is hypothesized in keeping with the findings by Fuchs and colleagues around performance feedback using CBM data, it will.
2. Is the DBR-CM perceived as usable within a performance feedback model, in terms of acceptability, understanding, feasibility, and system support?

Chapter 3

Method

Participants. Participants in this study included 16 regular education teachers in a rural, Midwestern elementary school serving students in third, fourth, and fifth grades. Of these 16 teachers, a subsample of five teachers was identified for participation in the performance feedback intervention (see Procedures section for further explanation). All participants in this study self-identified as Caucasian females. Participants reported years of experience ranging from less than one year to over 11 years (mean = 9). All participants were state certified elementary education teachers. Years of education ranged from a bachelor's degree to master's degree plus 30 hours of additional graduate credit hours ("Master's + 30"). Demographic information for teachers included in the performance feedback intervention is summarized in Table 2.

Teacher 1. Teacher 1 was a Caucasian female teaching in a 4th grade classroom. Teacher 1 was in her third year teaching at the time of this study. She was an elementary certified teacher by the state department of elementary and secondary education. No additional formal educational hours towards a master's degree or additional certifications were reported. She reported having 22 students in her classroom at the time of this study.

Teacher 2. Teacher 2 was a Caucasian female teaching in a 4th grade classroom. Teacher 2 was in her fourth year teaching at the time of this study. She was an elementary certified teacher by the state department of elementary and secondary education. She reported receiving a master's degree with a reading specialization. She reported having 20 students in her homeroom classroom at the time of this study.

Teacher 3. Teacher 3 was a Caucasian female teaching in a 4th grade classroom.

Teacher 3 was in her first year teaching at the time of this study. She was an elementary certified teacher by the state department of elementary and secondary education. No additional formal educational hours towards a master's degree or additional certifications were reported. She reported having 19 students in her homeroom classroom at the time of this study.

Teacher 4. Teacher 4 was a Caucasian female teaching in a 5th grade classroom. Teacher 4 was in her first year teaching at the time of this study. She was an elementary certified teacher by the state department of elementary and secondary education. No additional formal educational hours towards a master's degree or additional certifications were reported. She reported having 21 students in her homeroom classroom at the time of this study.

Teacher 5. Teacher 2 was a Caucasian female teaching in a 4th grade classroom. Teacher 2 was in her ninth year teaching at the time of this study. She was an elementary certified teacher by the state department of elementary and secondary education. She reported receiving a master's degree with a reading specialization, plus an additional 30 hours of formal graduate level training. She reported having 19 students in her homeroom classroom at the time of this study.

Measures. *Brief Classroom Interaction Observation-Revised.* The BCIO-R (Reinke & Newcomer, 2010) observations were completed by external observers using the Multi-Option Observation System for Experimental Studies (MOOSES; Tapp, 2002). MOOSES allows the use of handheld computers or tablets to collect data in real time; handheld computer devices were used in this study where each classroom was observed for 20 minutes. Frequency data was collected for teacher use of behavior specific praise,

general praise, precorrections, opportunities to respond, explicit reprimands and harsh reprimands simultaneously during each observation. Operational definitions for direct observation variables included in this study are included in Table 3. In addition, the duration of time a teacher spent actively teaching was recorded during each observation. All frequency data gathered using the BCIO-R were converted to rate per minute with the exception of the Time Teaching (maximum of 15 minutes). Time teaching was converted to the percent of time of the observation the teacher taught.

Reliability checks were conducted for all direct observations using the BCIO-R. A rating was considered reliable if multiple raters each indicated an occurrence of a targeted behavior no more than 5 seconds apart. If a match was found, then an agreement for that variable was tallied. Variables that were not matched were tallied as disagreements. An agreement ratio was then reported for each variable (agreements divided by the sum of agreements plus disagreements). The mean percentage agreement is reported as well as range of reliability for each variable.

Direct Behavior Rating - Classroom Management (DBR-CM). The DBR-CM was developed as an efficient measure of teacher classroom behavior. The DBR-CM utilizes a Likert scale rating system, ranging from 0 to 10 consistent with the DBR-SIS rating system, to rate each of five constructs of classroom management. Scores for each construct can be added together to compute an overall DBR-CM score. This study utilized the DBR-CM external rater (ER) form.

The operational definitions for each broad behavior domain are included on the DBR-CM form and are reported in Table 1. Briefly, praise is operationally defined as the use of positive praise statements and actions in response to the behavior and performance

of students in the classroom and a visibly general positive attitude towards all students. Communication is operationally defined as clearly conveying goals and expectations of a classroom to students. When applied to classroom management, rapport is operationally defined as is the quality of the student-teacher relationship, especially that of apparent mutual trust, emotional affinity, acceptance, and positivity. Engagement is operationally defined by the DBR-CM as 90% or more of students being involved in, focused on, or participating in instruction and/or classroom activities, and frequent presentation of questions that require a response to students by the teacher. Finally, Enthusiasm is operationally defined as the delivery of instructional content in a meaningful, memorable, and/or engaging manner. Specific behaviors, including examples and non-examples of each of the five constructs are included on the DBR-CM form. Inter-rater reliability reported in mean percent of agreement ranges from 67% to 81% (Praise = 69%, Communication = 81%, Engagement = 69%, Enthusiasm = 78%, and Rapport = 67%; Sims, Reinke, Riley-Tillman, & Herman, In Prep.).

It should be noted these data reflect inter-rater reliability for observations conducted in the absence of any formal training or reliability checks. Training and reliability checks had not been developed for use in the study to preliminarily examine the psychometric properties of the DBR-CM. For the present study, training and reliability checks had been developed and were used. Early psychometric examinations found DBR-CM scores were positively correlated with the overall CAS score ($r = .81$, $p = .01$) and the BCIO-R: Percent of Positive Implementation ($r = .53$, $p = .05$; Sims, Reinke, Riley-Tillman, & Herman, In Prep.).

User Rating Profile-Intervention (URP-I). To assess the usability of a performance feedback intervention using the DBR-CM, teachers completed the User Rating Profile-Intervention (URP-I, Chafouleas, Briesch, Riley-Tillman, & McCoach, 2009). The URP-I asks individuals to respond to a series of items examining the perceived usability of an assessment tool or intervention. Items are scored using ratings from 1 to 6. Lower scores indicate disagreement with statements in regard to intervention usability, while higher scores indicate agreement with these statements (e.g. 1 = *Strongly Disagree*, 2 = *Disagree*, 3 = *Slightly Disagree*, 4 = *Slightly Agree*, 5 = *Agree*, and 6 = *Strongly Agree*). To assess the usability of the DBR-teachers, observers, and data collectors that participated in the study will complete CM form, a modified URP-I. The revised URP-I, developed by Briesch and colleagues (2013), yielded a four-factor structure that provides information concerning intervention usage in the areas of acceptability, understanding, feasibility, and systems support. Acceptability refers to how fair, appropriate, and effective a rater perceives and intervention to be, including how enthusiastic she or he would be to use it. Items making up the factor labeled Understanding ask a respondent whether she or he feels that she or he understands how, and has the requisite skills, to carry out an intervention. The Feasibility factor assesses whether or not the respondent feels that implementing a described treatment is feasible, particularly in terms of intrusiveness (i.e., time requirements). Items making up the Systems Support factor ask implementers to rate their believe that external support would be needed in order to implement an intervention (Briesch, Chafouleas, Neugebauer, & Riley-Tillman, 2013).

Procedures. This study was proposed and subsequently approved by the

University of Missouri Institutional Review Boards (IRB). Informed consent was obtained from all participating teachers. Undergraduate and graduate students serving as data collectors for this study completed training and reliability checks for the DBR-CM and BCIO -R. Training included presentation of operational definitions, video examples, and non-examples of classroom management, discussion, and practice. Following this training, data collectors watched and rated example videos of teacher classroom management behavior. Data collector ratings were compared to master codes that were developed through an expert panel coding process to determine reliability. This process was repeated with MOOSES BCIO-R observation processes. Data collectors repeated training until they reached 85% agreement on each of the assessments. DBR-CM reliability was defined as concurrent ratings that fall within one point in either direction of the master code rating on each domain. BCIO -R reliability was defined as rating a behavior occurrence within 5 seconds of another rater's coding of that same behavior occurrence.

Following data collection training, each of the 16 trained teachers was observed by an external rater three times to form a baseline indication of classroom management skills. Real time measures of classroom management skills as measures by BCIO-R was collected during each observation session with the MOOSES system. In addition, at the end of the observation period, another external rater completed the DBR-CM. Next, a smaller subsample of five teachers was selected for the intervention phase of the study. This multiple-baseline, single case design evaluated the effectiveness of performance feedback using the DBR-CM. This subsample of teachers was chosen based on their median rate of overall praise score calculated using the combination of BCIO-R

variables. From the larger sample of 16 teachers, the teacher with the median overall rate of praise score, the two teachers immediately above the median, and the two teachers below the median were chosen for inclusion in the experimental component of this study. The middle performing teachers were identified as optimal participants in the intervention based on the hypothesis that performance feedback, in the absence of additional skill development training, is not designed to intervene in cases of severe classroom management skill deficit (e.g. the lowest DBR-CM ER total scores). Likewise, intervention with teachers with the highest classroom management performance scores may not require intervention or intervention of any kind may not result in noticeable change in classroom management performance.

Dependent variable. The identified outcome variable for this study was defined as the overall rate of praise. Overall rate of praise is calculated by combining BCIO-R variables to yield a score that represents overall positive classroom management. To calculate overall rate of praise, the rate of proactive classroom management strategies (e.g. praise and precorrection) were added together and divided by the total rate of observed practices (e.g. praise, precorrections, and reprimands) and multiplied by 100. The conversion of observed variables from the BCIO-R to a larger construct of classroom management has been used in prior studies (Reinke, Stormont, Herman, & Newcomer, 2014). The BCIO-R has been shown to be a reliable ($Kappa = .79$) and valid measure ($r = .53, p < .001; r = .52, p < .001$) of several discrete variable that make up classroom management (see Reinke, et al., 2015; Sims, Reinke, Riley-Tillman, & Herman, In Prep.).

Independent variable. The performance feedback intervention consisted of the

hand delivery of a line graph of DBR-CM scores (see Figure 1). This delivery followed each external observation during an individual teacher intervention phase. Feedback was hand-delivered to ensure that teachers view the graph and have the opportunity to ask clarification questions. Graphs included all DBR-CM scores. Broad behavior domains were color coded and labeled (e.g. Praise was a blue line, Communication was a red line, etc.). No specific qualitative performance feedback (e.g., coaching) was automatically included as part of this intervention. If teachers asked clarification questions, the delivery person answered questions as able.

To ensure accuracy of data collection observations, 30% of all data collection observations were conducted with multiple raters. Inter-observer agreement was calculated between BCIO-R data collectors by calculating the percentage of agreement in occurrence when raters score the same occurrence within 5 seconds of each other. DBR-CM reliability was calculated by comparing ratings on each construct, for a given observation period, and calculating the percentage of constructs that were rated within one point of each other. Inter-rater reliability will be calculated for BCIO-R and DBR-CM by comparing ratings between two external raters. If inter-rater reliability dropped below 85% for either data collection method, data collectors will receive additional training to improve reliability to at or above 85%. Across the study, inter-rater reliability for both the DBR-CM and BCIO-R variables exceeded 90%.

Baseline data collection continued and the five participating teachers' entry into their respective intervention phases were staggered over the next 40 days. Individual teacher entry into the intervention phase was initiated based on stability in baseline data for a given teacher. A stable pattern is defined as the identification of a trend in data so as

to provide an indication of a subsequent score in the absence of any treatment changes (T Chris Riley-Tillman & Burns, 2010). Stable baseline data and continued assessment through intervention phases allows for prediction, affirmation of the consequent, verification, and replication (by affirmation of the consequent; Riley-Tillman & Burns, 2010; Riley-Tillman & Walcott, 2007). Stable baseline data is critical to establishing causal effects through what Riley-Tillman and colleagues have termed “baseline logic” (Riley-Tillman & Burns, 2010; Riley-Tillman & Walcott, 2007). In baseline logic, as data is collected in the baseline phase, a pattern should emerge that allows SCD users to predict (e.g. Prediction) future performance if conditions are held constant. After a stable pattern was established in the external rater data for a teacher in the intervention phase (approximately one week), a new teacher with stable baseline data entered the intervention phase. This process continued until all teachers began receiving performance feedback via the DBR-CM. Decisions for changing phases was based on stability in the overall rate of praise BCIO-R variable, the targeted outcome variable for this study.

Originally, planned study procedures called for data collection to assess durability of behavior change following termination of formal intervention efforts. Following the completion of the intervention phase for the final teacher, observers were to return after four weeks to assess durability of intervention effects. This was not completed because of proximity the conclusion of intervention data collection to the end of the school year (approximately two weeks). At the conclusion of the final teacher’s intervention phase, each of the five teachers participating in the performance feedback intervention completed the URP-I to assess perceived usability of this performance feedback intervention using the DBR-CM.

Design. *Single Case Design.* Single Case Design (SCD) constitutes a prominent and practical methodology for educational researchers in particular to demonstrate functional relations between an identified behavior (e.g. dependent variable) and changes in practices or environmental factors (e.g. independent variables; Cipani, 2009; Levin, O'Donnell, & Kratochwill, 2003). SCDs are widely used to evaluate intervention effects in several educational applications including special education, developmental disabilities, behavioral disorders, and instructional strategies aimed at improving the performance of individual students (Shadish & Sullivan, 2011). SCD research methodology is characterized by the repeated measurement of the same behavior within a single experimental condition, implementation of each treatment or experimental condition across each participant, and the analysis of results within subject. This allows for each individual participant to serve as the basis of comparison for experimental effects (Cipani, 2009). This makes this approach particularly well-suited for educational applications as it allows all subjects to reap potential intervention effects. Repeated assessment is necessary due to the type of analysis conducted in SCD, representing the evaluation of within-subject changes rather than between-group comparisons. Group-design research, treatment effects are averaged across the entire group, which may result in individual differences in treatment effectiveness being lost.

A second advantageous characteristic of SCD is seen in the application of all treatments or interventions to all participants. In SCD, the experimental manipulation is applied to all participants, rather than specified individuals or groups of individuals, as is the case with group-design research. In SCD, the baseline data collection period is the equivalent of the control condition in group-design research. It is a measure of the target

behavior or skill before any formal treatment or instructional condition is imposed. Most SCDs utilize the condition to allow for a comparison of the rates of behavior before and after the designated intervention is implemented. After establishing stability within the baseline condition, participants enter the treatment phase where a specified change in practice is made. Depending on the selected SCD, the treatment may be added (e.g. AB design), staggered (e.g. multiple baseline design), removed and reimplemented (e.g. Reversal or ABAB design), followed by an additional treatment (e.g. Additive or ABC design), compared to several other treatments (e.g. multi-element or alternating treatment designs), or evaluated relative to changing criterion levels (e.g. changing criterion design).

Each of these designs seeks to replicate any apparent experimental effects. In SCDs, replication of experimental effects must be established both within and between studies. Replication of experimental effects is an essential part of scientific process, as replication establishes support for causal relationships between independent and dependent variables (Nock, Michel, & Photos, 2007). Effect replication in SCD research is an important mechanism for controlling threats to internal validity (Kratochwill et al., 2010). Horner and colleagues (Horner et al., 2005) advocate for a definition of replication as three demonstrations of the experimental effect at three different points in time with a single case (e.g. within-case replication), or across different cases (e.g. inter-case replication) within the design. Experimental effect is demonstrated when the dependent variable changes following manipulation of the independent variable to a remarkable degree (Kratochwill et al., 2010). This definition, three demonstrations of experimental effects, represents a conceptual norm in published articles, research, and textbooks that recommend methodological standards for single-case experimental designs (Kratochwill

et al., 2010). In this approach two important concepts must be considered, stability and trend in data. This approach is predicated on establishing a stable pattern of performance in the baseline phase and then analyzing changes from the pattern following introduction of some treatment designed to elicit change in targeted performance. Trending is increasing or decreasing data collected over time, a condition. Though trending in data may not be undesirable in treatment conditions, it can be problematic in baseline or no treatment conditions because if data is trending, it cannot also be stable. If data is trending in a desirable direction during the baseline condition, reliable conclusions about observed desirable effects in treatment conditions cannot be made. Stability or an undesirable trend (e.g. decreasing trend when the goal of a treatment is to increase performance) in data in the baseline condition, allows for any desirable treatment effects to be reliably attributed to the treatment. Stability of data is defined and demonstrated by a leveling off of data (Cooper, Heron, Heward, & others, 2007). Stability is important in SCD because it allows for some type of prediction to be made about what the data would have been if a treatment condition had not been introduced. Stability allows for change to be reliably assessed post treatment introduction. Unstable data within a condition prevents data collectors from predicting, with any confidence, a future rate of response under the same condition (Cipani, 2009). The number of data points required to establish stability within a condition is determined by the apparent variability of the behavior being measured (Cipani, 2009). If significant variability is noted in the data, additional data must be collected to ensure accurate predictions of future behavior can be made. If little variability is noted, less data collection is required.

Multiple-baseline design. The study employed a single-subject, across teachers,

multiple-baseline design, to determine the functional relationship between the independent variables (e.g. performance feedback delivered weekly) and the dependent variables (e.g. overall rate of praise). Multiple-baseline designs introduce the experimental condition to two or more subjects in temporal sequence (Nock et al., 2007). In a multiple-baseline design, there is one intervention condition following one baseline condition for each subject. In this design, the baseline phase is prolonged for the second and subsequent study subjects as entering the treatment phase is staggered for subjects. In this design, replication, the causal relationship, is established by demonstrating change occurs when, and only when, the intervention is directed at the dependent variable in question across multiple subjects (Nock et al., 2007). As discussed previously and consistent with other SCDs, multiple-baseline designs begin with a baseline phase that continues until behavioral stability is demonstrated for at least one subject, at which time the intervention condition begins (Cipani, 2009). When a performance change occurs only after an intervention is introduced, this suggests that the intervention caused the change. To rule out the influence of other factors that may account for the apparent change, in order to increase the validity causal claim, replication of the change must occur (Cipani, 2009; Riley-Tillman & Burns, 2010). Multiple baseline designs replicate the intervention with a new subject still in the baseline phase of the current study, rather than through a reversal or return to baseline approach (Cipani, 2009; Kazdin, 1982; Riley-Tillman & Burns, 2010). By establishing the same treatment results in performance change in multiple subjects in the same study when implementation is staggered over time and across subjects, researchers can reliably attribute these changes to a treatment.

Like other SCDs, multiple-baseline design offers many advantages over group-

designs as well as other SCD options (Cipani, 2009). Multiple-baseline design is flexible and can demonstrate causality by replicating treatment effects across different behaviors, different settings (e.g., school, home, and work), or different individuals (e.g., students, classrooms, teachers, or groups) (Nock et al., 2007). Arguably the best feature of SCD is the ability to apply the treatment to all participants. A major criticism of group-design is the withholding of treatment from a control group (Cipani, 2009; Kazdin, 1982; Levin et al., 2003). This is particularly advantageous in applied educational research. An additional benefit of multiple-baseline design is that it does not require removal of a treatment to establish replication of effect. Thus, once a subject is experiencing the benefits of an intervention, a multiple-baseline approach does not require the removal of the therapeutic effects of a treatment. Though the practice of intervention removal is valuable from a research perspective, many call attention to therapeutic and ethical concerns with the use of reversal designs (Nock et al., 2007). As noted previously, SCD allows for analysis of individual treatment effects, limiting the potential for such effects to be lost in the averaging of treatment effects across a group. This eliminates concerns regarding the reliability of data representing effects of experimental condition. SCD design offers another distinct advantage over the group-design in that it does not require large numbers of participants. SCDs can be conducted with as few as one participant, effectively eliminating the need for random sampling of participants.

The primary limitation of SCD is seen in weak generalizability of results. Multiple replications of studies are required to generalize findings beyond study participants. Additionally, multiple-baseline designs that demonstrate convincing causality require that change does not occur across behaviors, settings, or individuals

until the intervention is applied; therefore, interventions that lead to changes that generalize across all areas will limit the strength of the inferences that can be drawn (Nock et al., 2007).

Analytic plan. This study employed a multiple-baseline single-case design approach to evaluate the effect of performance feedback in the form of graphic presentation of DBR-CM scores on teacher classroom management. There are two main approaches to data analysis of single-case experimental designs: those involving visual inspection of graphed data and those using empirical statistical analyses (Kratochwill et al., 2010; Riley-Tillman & Burns, 2010). The two methods rely on the same methodological strategy to reach the goal. The goal of SCD focuses on establishing causal effects through what Riley-Tillman and colleagues have termed “baseline logic” (Riley-Tillman & Burns, 2010; Riley-Tillman & Walcott, 2007).

Visual analysis of data collected within SCD determines whether evidence of a relation between an independent variable and an outcome variable exists and the strength of the relation (Cipani, 2009; Kazdin, 1982; Nock et al., 2007; Riley-Tillman & Burns, 2010). Visual analysis for this study was conducted in a manner consistent with prevailing scholarly recommendations for SCD research (see (Kratochwill et al., 2010; Riley-Tillman & Burns, 2009). Within visual analysis, a causal relation can be inferred when a change in the outcome measure (e.g. dependent variable) appears to have followed the manipulation of a treatment (e.g. independent variable). A causal relation is demonstrated if the data across all phases of the study indicate at least three demonstrations of an effect at a minimum of three different points in time (Kratochwill et al., 2010). These analyses allow SCD users to evaluate the effect of an intervention by

comparing performance on the outcome of interest in a treatment-free condition to performance on the outcome of interest in a treatment condition. An effect is documented when the data pattern in one condition, treatment condition, differs more than would be expected from the stable data pattern observed in the previous condition (e.g., a baseline phase; Horner et al., 2005). Four primary criteria were proposed to guide visual inspection of single-case data by Kazdin (1982). First, the magnitude of the behavior change, or change in the mean and in level of the data collected in each phase, should be examined. Next, change in the rate of the behavior or an examination of differences in trend and latency to change should also be examined. A change in mean refers to the magnitude of change that occurs when the average score on the measure of the dependent variable in one condition is compared to another. A change in level describes immediacy, or lack thereof, of change in behavioral performance that occurs at the point between one condition and the next. This approach looks for clear, immediate change in level from one condition to the next. SCD research seeks to find large between-condition variability, which is only possible when minimal within-condition variability (stability) is present. In SCD, highly effective interventions would be evidenced by pronounced difference in between-condition trends and no latency to change between conditions.

Kratochwill and colleagues (2010) outline similar rules for conducting visual analysis of SCD data that uses four steps and six variables. Step one requires documentation of a predictable baseline pattern of data, or stability. Once a stable baseline pattern is documented, the second step consists of examining the data within each phase of the study to assess the within-phase pattern(s). For these steps to occur, sufficient data with sufficient consistency to demonstrate a predictable pattern of

responding must be present. The third step involves determining the effect of the independent variable on the dependent. This is accomplished through the visual analysis process comparing data from each phase with the data in the adjacent (or similar) phase to determine if an effect is present. If implementation of a treatment is associated with a predicted change in the pattern of the measurement data for the dependent variable, an effect is demonstrated. The fourth and final step in the visual analysis of SCD data involves integrating all the information from all phases of the study to determine whether there are at least three demonstrations of an effect at different points in time (e.g. documentation of a causal or functional relation; Kratochwill et al., 2010).

The What Works Clearinghouse Single-Case Design Technical Documentation for assessing effects within SCDs outlines six features used to examine within- and between-phase data patterns (Kratochwill et al., 2010). These features include level, trend, variability, immediacy of the effect, overlap, and consistency of data patterns across similar phases (Horner et al., 2005; Kazdin, 1982; Kratochwill et al., 2010). These six features, assessed individually and together, determine whether the results from a single-case study indicates a causal relationship between independent and dependent variables.

Ideally, with SCD visual inspection reveals intervention effects that are strong and clear making statistical analyses unnecessary. In instances where intervention effects are not clear, weaknesses in using visual inspection to evaluate interventions are evident. In these instances, it is necessary to examine the results more closely to determine if true effects are present that simply may be undetectable through visual analysis (Prentice & Miller, 1992). Another concern with visual analysis is the subjectivity and variability in

interpreting the results without the adoption of a clear and consistent threshold for demonstrating effectiveness. Visual analysis without specific decision rules allow for multiple interpretations of SCD data. This has led many to expand the evaluation of SCD data beyond visual analyses.

For this project, visual analysis was supplemented with empirical analyses utilizing effect-size estimates were calculated. In keeping with recommendations from Kratochwill (2010), several measures of effect size were calculated. outlined methods for examining effect size within SCD in response to identified issues. Percentage of Non-overlapping Data (PND), Percentage of All Non-overlapping Data (PAND), or Percent Exceeding the Median (PEM) are nonparametric methods recommended by What Works Clearinghouse for use in the analysis of SCDs (Kratochwill et al., 2010). Each of these methods provides a measure of the approximate size of the effect without formal statistical justification or distribution theory (Kratochwill et al., 2010).

Evaluating results. In a SCD multiple-baseline design, the experimental criterion is met by determining whether performance shifts following the introduction of the intervention. Riley-Tillman and Burns (2010) suggests that the experimental criterion for SCD studies is met by examining the effects of an intervention at different points in time. In order to evaluate intervention effects of performance feedback using the DBR-CM, overall rate of praise as determined by observation data was visually inspected for analysis of change in level, trend, variability, and immediacy/latency of effect (Riley-Tillman & Burns, 2010; Parsonson & Baer, 1978). To supplement visual analysis procedures, numerical estimates of level and trend for the phase means and slopes will be calculated for each measure.

Further analysis was conducted using the percentage of non-overlapping data points (PND) and the percentage of data points exceeding the median (PEM; Ma, 2006; Parker & Hagan-Burke, 2007). PND is thought of as the most straightforward, intuitive method with which to calculate effect size within SCD (Riley-Tillman, Methe, & Weegar, 2009; Scruggs, Mastropieri, & Casto, 1987). To address concerns with floor or ceiling limits in the measures used in this study PEM will be calculated (Ma, 2006; Parker, Hagan-Burke, & Vannest, 2007; Riley-Tillman et al., 2009). In addition, Tau-U statistic will be calculated as a measure of effect size (Parker & Vannest, 2011). Tau-U is a non-overlap statistic that can be used for to control for confounding baseline trend if needed, thus it is particularly well-suited for two-condition studies and small datasets that do not conform to typical parametric assumptions.

In order to assess the usability of the DBR-CM as a performance feedback intervention and as an assessment tool, individuals that completed the instrument will be asked to complete the URP-I to examine acceptability, understanding, feasibility, and systems support of performance feedback with the DBR-CM. High scores on the Acceptability subscale indicates a rater believes the intervention was appropriate given the problem behavior and would be personally enthusiastic about implementing the intervention. High on the Understanding subscale suggest the rater felt confident that she or he could implement the intervention given the description of the procedures. The higher the Feasibility subscale score, the higher the rater's confidence in his or her ability to implement the intervention with integrity given existing demands. A high Systems Support subscale score reflects the belief the intervention could not be implemented successfully without the assistance or support of other adults in the school environment.

Chapter 4

Results

Fidelity indices. Fidelity of implementation for this study was defined as delivery of DBR-CM results in a graph following each observation. Following each observation, DBR-CM scores for each of the four categories was entered into an excel spreadsheet. Scores were input into a line graph. Below each line graph, numeric values were listed and operational definitions of each DBR-CM broad behavior domain were included (see Figure 1). These graphs were then hand delivered to participants. Participants were asked to review their graphs and to ask any clarification question if they had them. Implementation fidelity for this process was 100%. On two occasions participants asked clarification question. After

Reliability checks were conducted for 27% of all observations. The mean percent agreement on the BCIO-R was 97% (90-100%). MOOSES utilizes a rigorous second-by-second comparison of raters to determine reliability and an overall reliability of 80% is considered acceptable (Tapp, 2002).

Visual analyses. *Teacher 1.* Visual analysis of BICO-R overall rate of praise data for Teacher 1 appears to indicate the most positive response to this intervention (see Figure 2). Overall, the level, or the apparent average score of data within a condition (Riley-Tillman & Burns, 2009), of data appears to be higher in the intervention phase when compared to baseline level. Most notably, baseline data indicate a negative trend, or the apparent slope of the best-fitting straight line for the data within a condition (Riley-Tillman & Burns, 2010), while intervention data show a positive trend. Visual analysis of variability, or the range or standard deviation of data around the trend line (Riley-Tillman

& Burns, 2009), between phases contributes little to the determination of intervention effects. Teacher 1 also showed immediacy (or latency), or apparent change in level between the last three data points in one phase and the first three data (Riley-Tillman & Burns, 2009), of intervention effects. The first three data points in the intervention phase appeared to increase from the last three data points in the baseline phase. Overall, visual analysis of data suggests some intervention effects

Teacher 2. Minimal intervention effect is noted in the visual analysis of plotted overall rate of praise data for Teacher 2 (see Figure 2). The level of data across phases for Teacher 2 does not appear to vary significantly. Similarly, baseline data indicates a flat to slightly negative trend, while intervention data appears to show a similar trend. No remarkable change in the variability of the data is noted between phases. Like Teacher 1, Teacher 2 showed immediate effects in response to intervention. Again, the first three data points in the intervention phase appeared to increase from the last three data points in the baseline phase. Despite some indication of intervention effectiveness in immediacy, overall visual analysis indicates the intervention was ineffective.

Teacher 3. Overall, visual analysis of data for Teacher 3 indicates some positive response following intervention implementation (see Figure 2). The level of data appears to increase slightly from baseline to intervention phase. Similarly, baseline data indicates a flat trend, while intervention data appears to show a positive trend. Variability appears to increase in intervention phase from baseline phase. No immediacy of intervention effectiveness is noted in the intervention phase data for Teacher 3.

Teacher 4. Overall, visual analysis of data for Teacher 4 indicates no intervention effects (see Figure 2). The apparent level of data across phases does not appear to vary

significantly across phases. Similarly, a flat trend is observed in both the baseline and intervention phase for Teacher 4. No remarkable change in the variability of the data is noted between phases. No immediate change in data is noted across phases for Teacher 4.

Teacher 5. Overall, visual analysis of data for Teacher 5 also indicates no intervention effects (see Figure 2). The apparent level of data across phases does not appear to vary significantly across phases. Similarly, a flat trend is observed in both the baseline and intervention phase for Teacher 4. No remarkable change in the variability of the data is noted between phases. No immediate change in data is noted across phases for Teacher 5.

Empirical analyses. To support visual analysis of data collected, advanced slope, mean, and non-overlap statistics were calculated for the identified outcome variable, BCIO-R: Overall Rate of Praise. First, the mean and slope of each phase, baseline and intervention, for each participant were calculated (see Table 4). Mean overall rate of praise scores improved from baseline to treatment for all participants. The mean score for Teacher 1 increase by 0.86 points. The mean score for Teacher 2 increase, 0.28 points. The mean score for Teacher 3 increase 0.18 points. The mean score for Teacher 4 increase 0.07 points. The mean score for Teacher 5 increase 0.05 points. The slope of the overall rate of praise improved from baseline to treatment for all but one participant. The difference in slope ranged from .005 to .07, with Teacher 1 showing the greatest improvement in rate of praise.

Further empirical analyses were conducted by calculating Percent of Non-overlapping Data (PND), Percent Exceeding Median (PEM), Tau-U, and Percent of All Non-overlapping Data (PAND) statistics. These data are presented in Table 5. PND,

PEM, and PAND were calculated using procedures outlined by Riley-Tillman and Burns (2010). Tau-U statistics were calculated using web-based calculator (Vannest, Parker, & Gonen, 2011). Scruggs and colleagues (1987) outline interpretation guidelines for PND statistics. These guidelines suggest PND scores of 90% or higher are consistent with very effective interventions, PND scores falling between 70% and 90% indicate effective interventions; PND scores that fall between 50% and 70% indicate interventions with questionable effect; and a PND score of 50% or lower indicate an intervention was ineffective. Calculated PND scores for teachers 1 through 5 are 43%, 20%, 21%, 0%, and 0%, respectively. Scruggs and colleagues (1986) offer the following interpretation guideline for PND: less than 50% reflects unreliable treatment, 50%-70% reflects questionable effectiveness, 70%-90% reflects fair effectiveness, and 90% and higher reflects high effectiveness.

Similar percentages are suggested for interpretation of PEM statistics. Scruggs and colleagues (Scruggs et al., 1987) suggest PEM scores at 90% or above indicate a high intervention effectiveness, intervention, scores between 70% and 90% are consistent with moderate intervention effectiveness, those scores between 50% and 70% are indicative of questionable intervention effectiveness, and a PEM score at 50% or lower indicates an ineffective intervention. PEM statistics for Teachers 1 through 5 are 43%, 90%, 50%, 0%, and 67% respectively. Parker and Vanest (2011) suggests a PEM value between 90 and 100% reflects highly effective treatment, 70 and 90% reflects moderately effective treatment, and less than 70% reflects questionable or not effective treatment.

Parker and Vannest (2012) provide guidance for interpreting Tau-U statistics. Tau-U scores range from 0% to 100% with score of 65% or lower indicating weak or

small intervention effects, score of between 66% and 92% indicating medium to high intervention effects; and scores above 93% indicating large or strong intervention effects. Tau-U statistics for Teachers 1 through 5 are 83%, 51%, 0%, 28%, and 27% respectively. Interpretation guidelines for Tau-U suggest a minimum value of 65% to establish minimal effectiveness of intervention (Parker, Vannest, & Davis, 2011).

Finally, PAND is reported as Cohen's d for interpretation (Parker, Hagan-Burke, & Vannest, 2007). Calculations of PAND and Cohen's d were completed by creating a 2 x 2 contingency table (see Table 6), in a manner consistent with the procedures outlined by Rile-Tillman and Burns (2010). Calculation of this statistic was determined appropriate given data across all subjects and phases was balanced (e.g. equal numbers of data points in baseline and intervention phases across subjects and phases). Analysis yielded a PAND of 36%. This was subsequently converted to Cohen's d to determine overall effect size of this intervention. This conversion resulted in a Cohen's d of -0.59. Suggested interpretation guidelines for Cohen's d indicated effect statistics below 0.2 are small, 0.5 are moderate, and 0.8 are large (Rosnow & Rosenthal, 1996).

Social validity. The results of teacher ratings on the URP- I yielded scores indicating a performance feedback intervention using the DBR-CM is minimally socially acceptable. The results of teacher responses on the URP-I are reported in Table 7. Acceptability scores ranged from 3 to 5 and resulted in a mean score of 4 (1 = *Strongly Disagree*, 2 = *Disagree*, 3 = *Slightly Disagree*, 4 = *Slightly Agree*, 5 = *Agree*, and 6 = *Strongly Agree*). Acceptability subscale measures a rater's enthusiasm about implementing an intervention. Understanding scores ranged from 3 to 6 and yielded a mean score across raters of 5. The understanding subscale measures rater's confidence in

implementation of the intervention given the description of the procedures. The mean Feasibility subscale score for all raters was a 4, with individual scores ranging from 3 to 5. This subscale measures the rater's confidence in his or her ability to implement the intervention with integrity given existing demands. Rater responses yielded a Systems Support subscale scores that range between 3 and 4. This resulted in a mean score of 3 across all raters. The Systems Support subscale reflects a rater's belief that the intervention could not be implemented successfully without the assistance or support of other adults in the school environment.

Chapter 5

Discussion

The DBR-CM was created to serve as an additional assessment option for teacher classroom management behavior. The DBR-CM assesses five core constructs of classroom management: praise, communication, rapport, enthusiasm, and engagement. It was hypothesized that performance feedback of teacher CM behaviors would have similar positive effects as previous performance feedback interventions (see Noell et al., 1997; Reinke et al., 2007; Sanetti et al., 2007; Scheeler et al., 2004). This study sought to determine if performance feedback, using the DBR-CM, would increase the classroom management behavior of participating teachers. Specifically, that brief exposure to these CM categories, and subsequent feedback on performance relative to these categories, would result in heightened awareness and behavioral change for teachers. To answer this question, a single case multiple-baseline design was used to determine performance feedback intervention effects on teacher CM behavior. Teachers were given a graphic representation of their DBR-CM scores following an observation conducted by an external observer. The feedback was staggered across five subjects over 45 days. Additional research questions sought to evaluate perceived usability of a performance feedback intervention targeting CM behavior using the DBR-CM.

Research Question 1: Does performance feedback, using the DBR-CM, increase the classroom management skills of participating teachers? Data collection across the five subjects included in the performance feedback intervention failed to demonstrate the minimum three replications of desired intervention effects required to establish intervention effectiveness (Kratochwill et al., 2012). Minimal intervention effects are

noted in the visual analysis in regard to trend and level of Teachers 1 and 3. Additional replications of these minimal intervention effects are not noted in the remaining subjects. Visual analysis of variability and immediacy of interventions are unremarkable. This failure to demonstrate intervention effectiveness across three subjects indicates the intervention in this form was unsuccessful in affecting positive change in teacher classroom management behavior as measured by overall rate of praise. These results suggest this performance feedback intervention, in this form, failed to replicate significant intervention effects across a minimum of three subjects.

Visual analysis of the data was followed by advanced empirical analyses employed in this study. Several non-overlap statistics were calculated. Percent of Non-overlapping Data (PND), Percent Exceeding Median (PEM), Tau-U, and Percent of All Non-overlapping Data (PAND) results are consistent with visual analytic findings. Though desirable intervention effects are evident for individual participating teachers (e.g. Teachers 1 & 3), these effects were limited to an individual statistic for a given teacher and replication of effects is not evident across teachers. For example, the PEM value for Teacher 3 is 90%, which indicates good intervention effects. Unfortunately, the PND and Tau-U values do not reach levels needed to establish intervention effects. The PEM value for Teacher 3 is the only value to reach a level that indicates an intervention effect. PAND statistics are also consistent with little to no intervention effect as the result of this performance feedback intervention.

Despite the results of this intervention indicating a performance feedback intervention using the DBR-CM was unsuccessful, positive effects of performance feedback for specific classroom management behaviors has been well documented (see

Reinke, Lewis-Palmer, & Martin, 2007, Reinke and Merrell, 2008). Several explanations for the absence of desirable intervention effects in this study are hypothesized, including poor performance of the assessment tool used to provide feedback (e.g. the DBR-CM), poor design of the performance feedback intervention, poor participant selection, and general teacher resistance.

Given the use of a novel assessment of classroom management in this study, a glaring explanation for the failure to induce positive intervention effects by this intervention is noted in the data yielded by the DBR-CM. Several aspects of the DBR-CM are open to scrutiny in this regard. As noted previously, prior research documents effectiveness of poor performance feedback in improving specific classroom management behaviors (e.g. Specific Praise). Similarly, research has established the utility of performance feedback to improve teacher treatment integrity (see DiGennaro, Martens, & McIntyre, 2005; Noell et al., 1997; Solomon et al., 2012). Within each of these approaches there may be a greater degree of specificity in the data collection and subsequent feedback. The DBR-CM, and its orientation as general outcomes measure, may lack the specificity needed to elicit positive intervention effects. Lack of intervention effects may also be explained by a limited utility in the DBR-CM data provided. This is to say that the DBR-CM data provided to participants may not provide useful information to teachers. In interpreting data, DBR-CM data may not have indicated a need for improvement. When combined, these issues may have resulted in the inability of teachers to discern what classroom management behaviors were in need of improvement. DBR-CM data may also be critically examined in regard to the reliability, validity, and interpretability of its assessment of classroom management. Given previously identified

acceptable convergent validity with other measures of classroom management, including specific SDO variables, poor quality, specificity, utility, and accuracy of DBR-CM data as the explanation for the intervention ineffectiveness is unlikely, but possible. Visual comparison of DBR-CM Praise scores to BCIO-R: Overall Rate of Praise calculated using BCIO-R variables. To facilitate this comparison, DBR-CM and BCIO-R: Overall Rate of Praise variables were converted to t-scores (Mean = 100, Standard Deviation = 15). Visual analysis indicate t-scores for both data sources were remarkably similar (see Figure 3). This similarity suggests the ability of the DBR-CM praise domain to measure praise used by teacher in classroom consistently and accurately in comparison to concurrent direct observations measures of praise. This further indicates the inability of this intervention to influence teacher classroom management behavior is likely not influenced by the DBR-CM assessment tool. However, still in question is the qualitative interpretability of DBR-DM data. To date, information indicating abnormal or at-risk classroom management behavior performance, as determined by DBR-CM score has not been established. When interpreting feedback as part of this intervention, teachers are left to determine subjectively if a score on any given broad behavior domain is good, bad, or somewhere in between. If this type of qualifying information were available to facilitate interpretation by participants, more desirable intervention effects may have resulted. Concerns around the quality, utility, accuracy, and interpretability of DBR-CM data as an explanation for intervention failure are reasonable, but unlikely given other potential reasons for observed intervention ineffectiveness.

A likely explanation for the ineffectiveness of the performance feedback intervention using the DBR-CM is seen in the design or structure of the intervention. As

noted previously, common iterations of performance feedback interventions are rooted in a SBC approach (Noell et al., 2005; Reinke et al., 2008). Such approaches involve more qualitative, coaching oriented performance feedback. This feedback frequently seeks to develop skill or motivation to change (Reinke, Herman, & Sprick, 2011). This intervention lacked such components. This may suggest that these components, rather than exposure to performance data alone, may be the mechanism that affects change in performance. In this intervention, teachers were given the option to solicit additional qualitative feedback, but chose to do so infrequently. It is possible that some participants may have wanted to improve their performance, but lacked the knowledge or skill to do so. In addition, participants may have felt uncomfortable asking for such qualitative feedback. Another potential flaw in the design of this intervention can be seen in the possible latency between delivery of performance feedback and opportunity to apply performance changes. Providing performance feedback following an intervention may create a large gap between feedback and opportunity to apply lessons learned from this feedback. This approach fails to tap into possible priming effects from delivery of feedback immediately prior to instructional periods in which desired changes in behavior could be used. Furthermore, it is possible that delivery of performance feedback using a line graph or with a graph of DBR-CM variables together may make interpretation by participants difficult. Delivery of broad behavior domains individually or in an alternative form (e.g. bar graph) may be easier for teachers to interpret. In short, within the design of this performance feedback intervention using the DBR-CM, several areas that may explain the resulting poor intervention effects are evident.

An additional explanation for the intervention ineffectiveness seen in this study

could be attributed to poor participant selection. The goal of this intervention was to target participants in need of reminders about classroom management rather than those in need of classroom management skill development. This goal was developed with a functional approach to behavior (Sugai et al., 2000) and Tier II, Response to Intervention principles in mind (D. Fuchs & Fuchs, 2006). In an attempt to reach this goal, teachers with the middle most overall praise rate scores were included in the treatment phase of this study. No normative information identifying classroom management behavior risk was used in this process. It is conceivable that the five selected participants were individuals with performance deficits that were inconsistent with the specified goal of including Tier II performance problems. It is possible that several or all of the participants were in need of more intensive interventions. Their need of a more intensive intervention paired with an intervention with a less than appropriate intensity, would have predisposed this intervention to failure.

An additional explanation for the apparent failure of this intervention to promote improvement in teacher classroom management behavior may be explained by general resistance to intervention efforts by teachers similar to those evident in school-based consultation. When teacher behavior change is either an indirect (e.g. consultation) or direct focus of intervention efforts, teachers often display resistance to these efforts for a variety of reason (Piersel & Gutkin, 1983). Intervention ineffectiveness in this study may be explained by fear of evaluation or criticism, time needed, self-confidence, distrust, poor problem solving skills, poor communication skills, attitude towards consultant, and other factors contributing to resistance identified by Gutkin & Hickman (1990) that were identified within SBC research. Though the simplicity of this intervention hoped to limit

the impact of such factors, there is likely no want to neutralize them completely. At the very least these factors may have limited participants' willingness to solicit additional qualitative feedback to aid in interpreting DBR-CM performance feedback.

Research Question 2: Is the DBR-CM perceived as usable within a performance feedback model, in terms of Acceptability, Understanding, Feasibility, and System

Support? At the conclusion of this study, participants were asked to provide information in regard to their perceptions of this performance feedback intervention targeting CM.

Data was collected using the URP- I. Findings suggest that participants agree somewhat with the overall utility of this intervention. Similarly, in each of the individual areas making up the URP- I, participants indicated they slightly agree or agree that this intervention is Acceptable, Understandable, Feasible, and incorporates appropriate Systemic Support. Overall, Teacher 3 expressed perceptions that this intervention was less usable than the other raters. Rater responses also suggest they slightly disagree that they received appropriate systemic support as part of implementing this intervention. In contrast, rater responses indicate higher perceptions in regard to understanding this intervention. In interpreting these findings, it should be noted that these perceptions are being reported for an intervention that had no effect. This suggests these finding have minimal practical utility, as the likelihood of replicating this intervention in its current form is small. With this in mind, participants indicated this intervention was not overly taxing and understandable. This information also indicates the potential for improvement in all areas assessed by the URP-I. Additional qualitative feedback obtained as part of the URP-I administration indicated participants believed that timing of this study within the school year as well as timing of external observations was problematic. Teacher 5

expressed frustration with the perceived delay in beginning her intervention phase. Several teachers also expressed feeling as if the external observations occurred at inopportune times. They expressed some frustration that observations occurred during student-directed or independent learning periods, which they felt limited their ability to engage in behaviors targeted by external observers (e.g. opportunities to respond, praise).

Limitations. Several limitations are evident within this study. First, it could be argued that the teachers included in the study resulted in a very homogenous sample. While the similarity in the subjects may decrease potential moderating factors, it also limits potential generalizability of finding. Next, a significant limitation of this study is noted in the timing of this study within the school year. Data collection occurred in the winter and spring of the school year. Care was taken to begin data collection at a point in the school year that maximized the likelihood of avoiding events like statewide high-stakes testing and spring break. Unfortunately, due to weather, scheduling issues inherent with applied research, and other less predictable school events, the intervention phase and associated data collection was influenced by these events. Intervention and associated data collection were interrupted by high-stakes testing, spring break, snow days, various class parties, field trips, and difficulty matching external observers and teacher instructional periods. Similarly, timing of external observations during the school day may be a limitation of this study. This potential limiting factor was echoed in the qualitative feedback provided by participants. An assumption of any classroom management assessment or intervention is that teachers are instructing or interacting with students to some degree. Several classroom observations for this study occurred during periods where teacher instruction or teacher-student interaction was limited, if not absent.

In such instances, both BCIO-R and DBR-CM scores appeared depressed. Due to an absence of instruction or interaction requiring CM, obtained scores may not accurately reflect teacher CM behavior. In some instances this was not the case, as observations during instruction or teacher-student interaction also yielded depressed CM scores, but timing of the observations should be considered in the future. Finally, a potential limitation of this study is noted in the latency for Teachers 4 and 5 between the beginning of baseline data collection and the onset of the intervention phase. Since this study occurred in the applied setting, additional subjects were included beyond those formally required by SCD standards in case of participant attrition. Visual and empirical analyses of between-phase changes for these subjects were particularly unremarkable. Given the timing issues for the study overall in conjunction with the delayed onset of performance feedback for these subjects relative to other participants, there is potential for these participant effects to have been negatively impacted.

Future directions. Though the results of this study indicate performance feedback using the DBR-CM was ineffective in producing increases in teacher CM behavior, the limitations noted earlier and the wealth of research supporting the use of performance feedback (see Noell et al., 1997; Reinke et al., 2007; Scheeler et al., 2004; Solomon et al., 2012) support replication of this study with refinements. In future performance feedback intervention applications of the DBR-CM, it may be beneficial to be more judicious in the selection of overall study timing as well as with timing of external observations. Implementing a performance feedback intervention using the DBR-CM at a point in the year less likely to coincide with other major calendar events such as holidays or testing may limit the influence these events may have had on the

results of this study. This in turn could lead to more desirable results. In addition, as part of a performance feedback intervention, taking care to ensure external observations are occurring during periods of teacher guided instruction or teacher-student interaction would limit potential depressive effects of classroom inactivity on scores. One possible solution is to replicate this study with a subject specific focus (i.e. Math teachers only, PBIS lesson instructional periods).

In future applications of the DBR-CM within performance feedback interventions, modifications to the procedures may result in more desirable outcomes. Potential modifications could include presentation of performance feedback prior to rather than following observations, adding a small but more formal training component to the intervention, and adding additional feedback components. Delivery of performance feedback prior to an observation period could result in a priming effect for teacher CM behavior. In this current iteration of a performance feedback intervention using DBR-CM, teachers were provided performance feedback after an observation period, which could mean they may not be required to engage in CM behaviors again until the following day. This latency between receiving performance feedback and opportunity to apply lesson learned from the performance feedback, may have resulted in the intervention ineffectiveness seen in this study.

Next, the addition of a small formal training component to the performance feedback intervention could influence intervention effectiveness. A small formal training around what constitutes performance feedback, in essence what external observers will be watching for, may support teacher understanding and use of good CM behavior. Another potential modification for future studies of performance feedback interventions

incorporating the DBR-CM is modifying the operational definition of performance feedback. In this study performance feedback was very simple. Performance feedback constituted providing graphic representation of DBR-CM scores. Teachers were given the opportunity to ask for clarification, but no explanation of the scores was automatically provided. By expanding the operational definition of performance feedback to include a component where the delivery of performance feedback includes interpretation or elaboration, intervention effects could be positively impacted. The specifics of these additional components would need to be planned carefully. The goal of this and future work with performance feedback interventions and the DBR-CM is to minimize resources and time associated with an intervention while maximizing positive outcomes consistent with a Tier II level intervention. There are several preexisting performance feedback-based interventions that are more time and resource intensive that have documented positive impact on teacher CM behavior (see Reinke et al., 2008; Reinke, Stormont, Webster-Stratton, Newcomer, & Herman, 2012; Sprick, 2013). The goal of this work is to stay as far away as possible from other very useful, very impactful coaching- or consultation-based performance feedback interventions. Though it is hoped that the DBR-CM could prove useful in these performance feedback intervention applications as well. Though the goal of this work was to explore the utility of a performance feedback intervention free of any formal consultative process, a future direction of this intervention applications of the DBR-CM its utility within such formats could be explored. Given the limited success of a performance feedback intervention in this format, it may be beneficial to explore use of the DBR-CM in more resource intensive formats. Though theoretically a format incorporating formal consultation processes may be more Tier III

oriented, such applications are no less important in a larger picture tiered approach to classroom management interventions.

A final future direction for this work is seen in replication of this study with fewer subjects. This study included 5 subjects and results failed to achieve replication of intervention effects across three subjects. Feedback provided by participants indicated that the onset of intervention for participants 4 and 5 may have created frustration (in combination with other delays experienced in the study). A study limited to 3 participants would eliminate the latency effects for any participants that would fall after the third participant. This would limit the frustration on the part of these teachers and potentially increase intervention effectiveness.

Conclusion. The present study examined a simplified performance feedback intervention targeting teacher classroom management behavior. This intervention was based on prior research suggesting the academic achievement increase with regular feedback in the absence of any formal intervention efforts (e.g. no change in resource allocation or instructional practices; Fuchs et al., 1991). This approach represented a less time and resource taxing approach to intervention for teacher classroom management behavior. Consistent with the time and resources needed (e.g. Tier II level intervention) to implement this basic performance feedback intervention, teachers displaying mild to moderate deficits in classroom management performance (e.g. Tier II level deficits) were targeted for inclusion in this study. Consistent with previous research, classroom management deficits were identified using measurement of overall use of praise in their classroom (Reinke et al., 2007; Stormont & Reinke, 2009). Performance feedback was delivered through graphic representation of DBR-CM assessment results. The DBR-CM

is a novel classroom management assessment incorporating both general outcomes measures (Shinn & Shinn, 2002) and DBR (Chafouleas, Riley-Tillman, & Christ, 2009) assessment methodologies. Immediately following an observation conducted by an external observer, DBR-CM scores were plotted on a line graph and delivered to participating teachers. Data was collected as part of this multiple-baseline single case design study was analyzed in a manner consistent with prevailing guidelines (Kratochwill et al., 2010; Riley-Tillman & Burns, 2010). Visual and empirical analyses of results indicated noteworthy improvements in overall teacher classroom management behavior in two of five included teachers. These findings indicate that a performance feedback classroom management intervention in this form failed to document the required three replication of intervention effects to be considered efficacious. Several factors may explain these findings including study timing, inappropriate inclusion of participants, teacher resistance, shortcomings in performance feedback information provided, and shortcoming of performance feedback dissemination method. Though this study failed to meet the single-case design standards to be considered efficacious, the positive findings in two participants and prior research suggests that continued exploration of this topic is warranted. Through refinement in both the screening and intervention processes, a less time and resources intensive performance feedback-based intervention that meets single-case design efficacy standards may be identified.

References

- Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology, 51*(1), 1–18.
- Barrett, S. B., Bradshaw, C. P., & Lewis-Palmer, T. (2008). Maryland Statewide PBIS Initiative Systems, Evaluation, and Next Steps. *Journal of Positive Behavior Interventions, 10*(2), 105–114.
- Bergan, J. R. (1977). *Behavioral consultation*. Columbus, Ohio: Merrill.
- Briesch, A. M., Chafouleas, S. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2013). Assessing influences on intervention implementation: Revision of the usage rating profile-intervention. *Journal of School Psychology, 51*(1), 81–96.
- Brophy, J. (1985). Teachers' expectations, motives and goals for working with problem students. *Research on Motivation in Education, 2*, 175–213.
- Brophy, J., & McCaslin, M. (1992). Teachers' reports of how they perceive and cope with problem students. *The Elementary School Journal*.
- Bruer, J. T. (1993). *Schools for thought: A science of learning in the classroom*. MIT press.
- Chafouleas, S. M. (2011). Direct Behavior Rating: A Review of the Issues and Research in Its Development. *Education & Treatment of Children, 34*(4), 575–591.
- Chafouleas, S. M., Briesch, A. M., Riley-Tillman, T. C., & McCoach, D. B. (2009). Moving beyond assessment of treatment acceptability: An examination of the

- factor structure of the Usage Rating Profile—Intervention (URP-I). *School Psychology Quarterly*, 24(1), 36.
- Chafouleas, S. M., Riley-Tillman, T. C., & Christ, T. J. (2009a). Direct Behavior Rating (DBR). *Assessment for Effective Intervention*, 34(4), 195–200.
- Chafouleas, S. M., Riley-Tillman, T. C., & Christ, T. J. (2009b). Direct Behavior Rating (DBR): An emerging method for assessing social behavior within a tiered intervention system. *Assessment for Effective Intervention*.
- Chafouleas, S. M., Riley-Tillman, T. C., & McDougal, J. L. (2002). Good, bad, or in-between: How does the daily behavior report card rate? *Psychology in the Schools*, 39(2), 157–169.
- Chafouleas, S. M., Riley-Tillman, T. C., & Sugai, G. M. (2007). *School-based behavioral assessment: Informing intervention and instruction*. Guilford Press.
- Christ, T. J., Riley-Tillman, T. C., & Chafouleas, S. M. (2009). Foundation for the development and use of direct behavior rating (DBR) to assess and evaluate student behavior. *Assessment for Effective Intervention*, 34(4), 201–213.
- Christ, T. J., Zopluoglu, C., Long, J. D., & Monaghan, B. D. (2012). Curriculum-based measurement of oral reading: Quality of progress monitoring outcomes. *Exceptional Children*, 78(3), 356–373.
- Christ, T. J., Zopluoglu, C., Monaghan, B. D., & Van Norman, E. R. (2013). Curriculum-based measurement of oral reading: Multi-study evaluation of schedule, duration, and dataset quality on progress monitoring outcomes. *Journal of School Psychology*, 51(1), 19–57.

- Cipani, E. (2009). *Practical research methods of educators*. New York: Springer Publishing Company.
- Cochrane, W. S., & Laux, J. M. (2008). A survey investigating school psychologists' measurement of treatment integrity in school-based interventions and their beliefs about its importance. *Psychology in the Schools, 45*(6), 499–507.
- Cone, J. D. (1978). The behavioral assessment grid (BAG): A conceptual framework and a taxonomy. *Behavior Therapy, 9*(5), 882–888.
- Cooper, J. O., Heron, T. E., Heward, W. L., & others. (2007). *Applied behavior analysis*.
- Corcoran, T. B. (1995). *Helping teachers teach well: Transforming professional development*. Consortium for Policy Research in Education New Brunswick, NJ.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clinical Psychology Review, 18*(1), 23–45.
- Deno, S. L. (1995). School psychologist as problem solver. *Best Practices in School Psychology III, 471–484*.
- Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37*(3), 184–192.
- DiGennaro, F. D., Martens, B. K., & McIntyre, L. L. (2005). Increasing treatment integrity through negative reinforcement: Effects on teacher and student behavior. *School Psychology Review, 34*(2), 220.
- DuFour, R., & Eaker, R. (1998). *Professional learning communities at work: Best practices for enhancing student achievement*. ERIC.

- Elmore, R. F. (2002). Bridging the gap between standards and achievement. *Albert Shanker Institute, (Washington, DC, 2002)*, 17.
- Elmore, R. F. (2005). Accountable leadership (Vol. 69, pp. 134–142). Presented at the The Educational Forum, Taylor & Francis.
- Erchul, W. P., & Ward, C. S. (2016). Problem-Solving Consultation. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of Response to Intervention* (pp. 73–86). Springer US.
- Feiman-Nemser, S. (1996). Teacher Mentoring: A Critical Review. ERIC Digest.
- Florian, L. (2007). Reimagining special education. *The Sage Handbook of Special Education*, 7–20.
- Fuchs, D., & Fuchs, L. S. (1993). Inclusive schools movement and the radicalization of special education reform.
- Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, 41(1), 93–99.
- Fuchs, L. S., Hamlett, D. F. C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal*, 28(3), 617–641.
- Gettinger, M., & Seibert, J. K. (2002). Best practices in increasing academic learning time. *Best Practices in School Psychology IV, 1*, 773–787.
- Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review*.

- Gresham, F. M. (2008). Best practices in diagnosis in a multitier problem-solving approach. *Best Practices in School Psychology V, 2*, 281–294.
- Gresham, F. M. (2008). Best practices in school psychology. In *Best practices in diagnosis in a multitier problem-solving approach*. (Vol. 2, pp. 281–294).
- Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice, 15*(4), 198–205.
- Guskey, T. R. (2000). *Evaluating professional development*. Corwin Press.
- Gutkin, T. B., & Hickman, J. A. (1990). The Relationship of Consultant, Consultee, and Organizational Characteristics to Consultee Resistance to School-Based Consultation: An Empirical Analysis. *Journal of Educational & Psychological Consultation, 1*(2), 111.
- Hagermoser Sanetti, L. M., & Fallon, L. M. (2011). Treatment integrity assessment: How estimates of adherence, quality, and exposure influence interpretation of implementation. *Journal of Educational and Psychological Consultation, 21*(3), 209–232.
- Hamre, B. K., & Pianta, R. C. (2001). Early teacher–child relationships and the trajectory of children’s school outcomes through eighth grade. *Child Development, 72*(2), 625–638.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

- Hintze, J. M., Volpe, R. J., & Shapiro, E. S. (2002). Best practices in the systematic direct observation of student behavior. *Best Practices in School Psychology, 4*, 993–1006.
- Horner, R. H. (2000). Positive behavior supports. *Focus on Autism and Other Developmental Disabilities, 15*(2), 97–105.
- 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.
- Huling, L.-R., & others. (2001). Teacher Mentoring as Professional Development. ERIC Digest.
- Kazdin, A. E. (1982). Single-case experimental designs in clinical research and practice. *New Directions for Methodology of Social & Behavioral Science*.
- Kratochwill, T. R., Alper, S., & Cancelli, A. A. (1980). Nondiscriminatory assessment: perspectives in psychology and education. *The Fourth Review of Special Education, 229–286*.
- Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2012). Single-case intervention research design standards. *Remedial and Special Education, 0741932512452794*.
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. *What Works Clearinghouse*.
- Lane, K. L., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2004). Treatment Integrity: An Essential—But Often Forgotten—Component of School-Based

- Interventions. *Preventing School Failure: Alternative Education for Children and Youth*, 48(3), 36–43. <http://doi.org/10.3200/PSFL.48.3.36-43>
- Levin, J. R., O'Donnell, A. M., & Kratochwill, T. R. (2003). Educational/psychological intervention research. *Handbook of Psychology*.
- Marzano, J. S., & Pickering, D. (2003). *Classroom management that works: Research-based strategies for every teacher*. ASCD.
- May, S., Ard, W., Todd, A. W., Horner, R. H., Glasgow, A., Sugai, G., & Sprague, J. (2000). School-wide information system. *Educational and Community Supports, University of Oregon, Eugene, Oregon*.
- McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review*, 33(2), 193.
- McKinney, C. W., Larkins, A. G., Kazelskis, R., Ford, M. J., Allen, J. A., & Davis, J. C. (1983). Some effects of teacher enthusiasm on student achievement in fourth grade social studies. *The Journal of Educational Research*, 249–253.
- Merrell, K. W. (2001). Assessment of children's social skills: Recent developments, best practices, and new directions. *Exceptionality*, 9(1-2), 3–18.
- Murnane, R., & Cohen, D. (1986). Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive. *Harvard Educational Review*, 56(1), 1–18.
- 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.

- Nock, M. K., Michel, B. D., & Photos, V. I. (2007). Single-case research designs. *Handbook of Research Methods in Abnormal and Clinical Psychology*, 337–350.
- Noell, G. H., Witt, J. C., Gilbertson, D. N., Ranier, 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.
- Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L., ... Duhon, G. J. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review*, 34(1).
- Oliver, R. M., & Reschly, D. J. (2010). Special Education Teacher Preparation in Classroom Management: Implications for Students with Emotional and Behavioral Disorders. *Behavioral Disorders*, 35(3), 188–199.
- Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of All Non-Overlapping Data (PAND) An Alternative to PND. *The Journal of Special Education*, 40(4), 194–204.
- Parker, R. I., & Vannest, K. J. (2012). Bottom-up analysis of single-case research designs. *Journal of Behavioral Education*, 21(3), 254–265.
- Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, 35(4), 303–322.
- Peterson, K. (2004). Research on school teacher evaluation. *NASSP Bulletin*, 88(639), 60–79.

- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). Classroom assessment scoring system. *Baltimore: Paul H. Brookes.*
- Pianta, R. C., & Stuhlman, M. W. (2004). Teacher-child relationships and children's success in the first years of school. *School Psychology Review, 33*(3), 444.
- Piersel, W. C., & Gutkin, T. B. (1983). Resistance to school-based consultation: A behavioral analysis of the problem. *Psychology in the Schools, 20*(3), 311–320.
- Pindiprolu, S. S., Peterson, S. M., & Bergloff, H. (2007). School Personnel's Professional Development Needs and Skill Level with Functional Behavior Assessments in Ten Midwestern States in the United States: Analysis and Issues. *JOURNAL-INTERNATIONAL ASSOCIATION OF SPECIAL EDUCATION, 8*(1), 31.
- Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin, 112*(1), 160.
- Reese, J. (2007). The four Cs of successful classroom management. *Music Educators Journal, 94*(1), 24–29.
- Reinke, W. M., Herman, K. C., & Sprick, R. (2011). *Motivational interviewing for effective classroom management: The classroom check-up.* Guilford Press.
- Reinke, W. M., Herman, K. C., & Stormont, M. (2013). Classroom-Level Positive Behavior Supports in Schools Implementing SW-PBIS Identifying Areas for Enhancement. *Journal of Positive Behavior Interventions, 15*(1), 39–50.
- Reinke, W. M., Herman, K. C., Stormont, M., Newcomer, L., & David, K. (2013). Illustrating the multiple facets and levels of fidelity of implementation to a teacher classroom management intervention. *Administration and Policy in Mental Health and Mental Health Services Research, 40*(6), 494–506.

- Reinke, W. M., Lewis-Palmer, T., & Martin, E. (2007). The effect of visual performance feedback on teacher use of behavior-specific praise. *Behavior Modification, 31*(3), 247–263.
- Reinke, W. M., Lewis-Palmer, T., & Merrell, K. (2008). The classroom check-up: A classwide teacher consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review, 37*(3), 315.
- Reinke, W. M., Stormont, M., Herman, K. C., Wang, Z., Newcomer, L., & King, K. (2014). Use of Coaching and Behavior Support Planning for Students With Disruptive Behavior Within a Universal Classroom Management Program. *Journal of Emotional and Behavioral Disorders, 1063426613519820*.
- Reinke, W. M., Stormont, M., Webster-Stratton, C., Newcomer, L. L., & Herman, K. C. (2012). The incredible years teacher classroom management program: Using coaching to support generalization to real-world classroom settings. *Psychology in the Schools, 49*(5), 416–428.
- Riley-Tillman, T. C., & Burns, M. K. (2009). Single case design for measuring response to educational intervention. *New York: Guilford*.
- Riley-Tillman, T. C., & Burns, M. K. (2010). *Evaluating educational interventions: Single-case design for measuring response to intervention*. Guilford press.
- Riley-Tillman, T. C., Chafouleas, S. M., & Briesch, A. M. (2007). A school practitioner's guide to using daily behavior report cards to monitor student behavior. *Psychology in the Schools, 44*(1), 77–89.
- Riley-Tillman, T. C., Chafouleas, S. M., Briesch, A. M., & Eckert, T. L. (2008). Daily behavior report cards and systematic direct observation: An investigation of the

- acceptability, reported training and use, and decision reliability among school psychologists. *Journal of Behavioral Education, 17*(4), 313–327.
- Riley-Tillman, T. C., Chafouleas, S. M., Christ, T., Briesch, A. M., & LeBel, T. J. (2009). The impact of item wording and behavioral specificity on the accuracy of direct behavior ratings (DBRs). *School Psychology Quarterly, 24*(1), 1.
- Riley-Tillman, T. C., Christ, T. J., Chafouleas, S. M., & Boice-Mallach, C. H. (2011). The impact of observation duration on the accuracy of data obtained from direct behavior rating (DBR). *Journal of Positive Behavior Interventions, 13*(2), 119–128.
- Riley-Tillman, T. C., Kalberer, S. M., & Chafouleas, S. M. (2005). Selecting the right tool for the job: A review of behavior monitoring tools used to assess student response to intervention. *The California School Psychologist, 10*(1), 81–91.
- Riley-Tillman, T. C., & Walcott, C. M. (2007). Using baseline logic to maximize the value of educational interventions. In *School Psychology Forum* (Vol. 1, pp. 87–97).
- Rosnow, R. L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Psychological Methods, 1*(4), 331.
- Salvia, J., Ysseldyke, J. E., & Bolt, S. (2007). *Assessment in special and inclusive education*. Wadsworth.
- Sanetti, L. M. H., Luiselli, J. K., & Handler, M. W. (2007). Effects of verbal and graphic performance feedback on behavior support plan implementation in a public elementary school. *Behavior Modification, 31*(4), 454–465.

- Scheeler, M. C., Ruhl, K. L., & McAfee, J. K. (2004). Providing Performance Feedback to Teachers: A Review. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, 27(4), 396–407. <http://doi.org/10.1177/088840640402700407>
- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research methodology and validation. *Remedial and Special Education*, 8(2), 24–33.
- Shadish, W. R., & Sullivan, K. J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods*, 43(4), 971–980.
- Sheridan, S. M., Welch, M., & Orme, S. F. (1996). Is consultation effective? A review of outcome research. *Remedial and Special Education*, 17(6), 341–354.
- Shinn, M. R. (2002). Best practices in using curriculum-based measurement in a problem-solving model. *Best Practices in School Psychology IV, 1*, 671–697.
- Shinn, M. R., & Shinn, M. M. (2002). AIMSweb training workbook: Administration and scoring of reading maze for use in general outcome measurement. *Eden Prairie, MN: Edformation*.
- Shinn, M. R., Shinn, M. M., Hamilton, C., & Clarke, B. (2002). Using curriculum-based measurement in general education classrooms to promote reading success. *Interventions for Academic and Behavior Problems II: Preventive and Remedial Approaches*, 113–142.
- Showers, B. (1985). Teachers coaching teachers. *Educational Leadership*, 42(7), 43–48.

- 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*(3), 351–380.
- Simonsen, B., Fairbanks, S., Briesch, A., & Sugai, G. (2006). Classroom management: Self-assessment revised. *Storrs, CT: OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports*. Retrieved June, 24, 2011.
- Simonsen, B., Myers, D., & DeLuca, C. (2010). Teaching teachers to use prompts, opportunities to respond, and specific praise. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children, 33*(4), 300–318.
- Sims Jr, H. P., & Manz, C. C. (1982). Social learning theory: The role of modeling in the exercise of leadership. *Journal of Organizational Behavior Management, 3*(4), 55–63.
- Solomon, B. G., Klein, S. A., & Politylo, B. C. (2012). The Effect of Performance Feedback on Teachers' Treatment Integrity: A Meta-Analysis of the Single-Case Literature. *School Psychology Review, 41*(2).
- Sprick, R. S. (2013). *Discipline in the secondary classroom: A positive approach to behavior management*. John Wiley & Sons.
- Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of Educational Change, 7*(4), 221–258.

- Stormont, M., & Reinke, W. (2009). The Importance of Precorrective Statements and Behavior-Specific Praise and Strategies to Increase Their Use. *Beyond Behavior, 18*(3), 26–32.
- Sugai, G., & Horner, R. (2002). The evolution of discipline practices: School-wide positive behavior supports. *Child & Family Behavior Therapy, 24*(1-2), 23–50.
- Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., ... Turnbull, A. P. (2000). Applying positive behavior support and functional behavioral assessment in schools. *Journal of Positive Behavior Interventions, 2*(3), 131–143.
- Tapp, J. (2002). *Multiple option observation system for experimental studies (MOOSES)[Software]*.
- Tapp, J., & Wehby, J. (1993). MOOSES: Multiple Option Observation System for Experimental Studies.
- Tapp, J., Wehby, J., & Ellis, D. (1995). A multiple option observation system for experimental studies: MOOSES. *Behavior Research Methods, Instruments, & Computers, 27*(1), 25–31.
- Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach*. Guilford Press.
- Tillery, A. D., Varjas, K., Meyers, J., & Collins, A. S. (2010). General education teachers' perceptions of behavior management and intervention strategies. *Journal of Positive Behavior Interventions, 12*(2), 86–102.
- Timperley, H. (2008). Teacher professional learning and development.

- Truscott, S. D., Kreskey, D., Bolling, M., Psimas, L., Graybill, E., Albritton, K., & Schwartz, A. (2012). Creating consultee change: A theory-based approach to learning and behavioral change processes in school-based consultation. *Consulting Psychology Journal: Practice and Research, 64*(1), 63.
- Wagner, M., Friend, M., Bursuck, W. D., Kutash, K., Duchnowski, A. J., Sumi, W. C., & Epstein, M. H. (2006). Educating Students With Emotional Disturbances A National Perspective on School Programs and Services. *Journal of Emotional and Behavioral Disorders, 14*(1), 12–30.
- Williams, J. (2005). Deleuze and Whitehead: the concept of reciprocal determination. *Cloots, A., and Robinson, K., Brussels (eds) Deleuze, Whitehead and the Transformation of Metaphysics, Brussels: Koninklijke Vlaamse Academie Van België Voor Wetenschappen En Kusten, 89–105.*
- Yell, M. L., Shriener, J. G., & Katsiyannis, A. (2006). Individuals with disabilities education improvement act of 2004 and IDEA regulations of 2006: Implications for educators, administrators, and teacher trainers. *Focus on Exceptional Children, 39*(1), 1–24.

Table 1.

Operational Definitions of Direct Behavior Rating-Classroom Management

Direct Behavior Rating-Classroom Management External Rater Form	
Praise	The use of positive praise statements in response to the behavior and performance of students in the classroom. Examples: Teacher uses more behavior-specific praise than general praise, uses praise contingent on expected behavior, provides three (3) or more praise statements for every reprimand, displays more positive than negative attitude when interacting with students, provides praise at desirable rates using non-verbal interactions such as gestures, tangibles, or physical contact, and maintains an overall tone that is positive and not negative or sarcastic.
Communication	The clear communication of goals and expectations of an instructional period. Examples: Teacher provides clear academic and behavioral expectations to the students, explicitly states or posts instructional objectives and offers opportunity for clarifying questions, clearly presents behavioral expectations verbally and/or visually, uses an attention signal to gain attention of all students, and utilizes transition procedures that appear to be known and followed by majority of students (as evidenced by efficient classroom transitions).
Engagement	90% or more of students engaged 80% of the time during instruction and/or classroom activities and students are provided and respond to questions posed to the group and individual students occur frequently. Examples: 90% of students are clearly academically engaged at all times; level of observable disruptions in the classroom is minimal; teacher provides four (4) or more opportunities for students to respond per minute during instruction; and teacher asks many different students in the classroom at least one question during instruction.
Enthusiasm	The delivery of instructional content in a meaningful, memorable, and/or engaging manner. Examples: Teacher tone and pace of instruction are positive and upbeat, instructional content is supplemented with or related to a familiar life application, topic, or activity, and instruction incorporates alternative activities (e.g. students as teachers, group work, pair and share, current event, etc.).
Rapport	The quality of the student-teacher relationship, especially that of mutual trust, emotional affinity, acceptance, and positivity. Examples: The general feel of the classroom is mutually warm and accepting; the teacher uses children's names frequently; interactions between the

teacher and students are visibly positive; students appear to feel comfortable approaching the teacher or asking questions; and teacher appears to feel comfortable, positive, and genuine in his/her interactions with students.

Table 2.

Pertinent Participant Demographic Data

	Gender	Race/Ethnicity	Education Level	Years of Experience
Teacher 1	F	Caucasian	Bachelors	2.5 years
Teacher 2	F	Caucasian	Masters	4.5 years
Teacher 3	F	Caucasian	Bachelors	>1 year
Teacher 4	F	Caucasian	Bachelors	2.5 years
Teacher 5	F	Caucasian	Masters +30	9 years

Table 3.

Operational Definitions of Direct Observation Variables

Teacher Frequency Codes	
Specific Praise	Verbal statement or gesture that indicates approval and <u>names a specific behavior</u> .
General Praise	Verbal statement or gesture that indicates approval <u>and does not name a specific behavior</u> .
Explicit Reprimand	Verbal comments or gestures by teacher to indicate disapproval of behavior; reprimand is concise (brief) in a normal speaking tone.
Harsh Reprimand	Verbal comments or gestures indicate disapproval of behavior using a voice louder than typical for setting or harsh, critical or sarcastic tone.
Opportunity to Respond	Instructional prompt (statement gesture, or visual cue) that requires immediate academic response to teacher.
Precorrective Statement	Teacher provides specific statement to prompt expected student behavior(s) before the behavior occurs. (e.g. before transition). The teacher must have anticipated potential for problem behavior and make the statement before problem behaviors occur. Direction prompts specific behavior expectation, not academic or content related tasks. [Wait until the end of the statement]
Student Frequency Codes	
Disruptive Behavior	Any statement or action by an individual student or group of students that interferes with ongoing classroom activities for the teacher and/or one or more peers.
Aggressive Behavior	Student is physically or verbally aggressive toward objects, peer(s) or teacher. If physical and verbal aggression occur simultaneously code for both (2 aggressive behaviors). You must observe the behavior to code it (do not rely on peer reporting aggression to teacher).
Duration code: Switch following observing the behavior for 5 seconds.	
Teaching 5-second rule	Teacher is engaged in instruction, active supervision, or is monitoring students as they work.
Not Teaching	Teacher is not engaging students and is involved in independent task with

5-second rule no interaction with students.

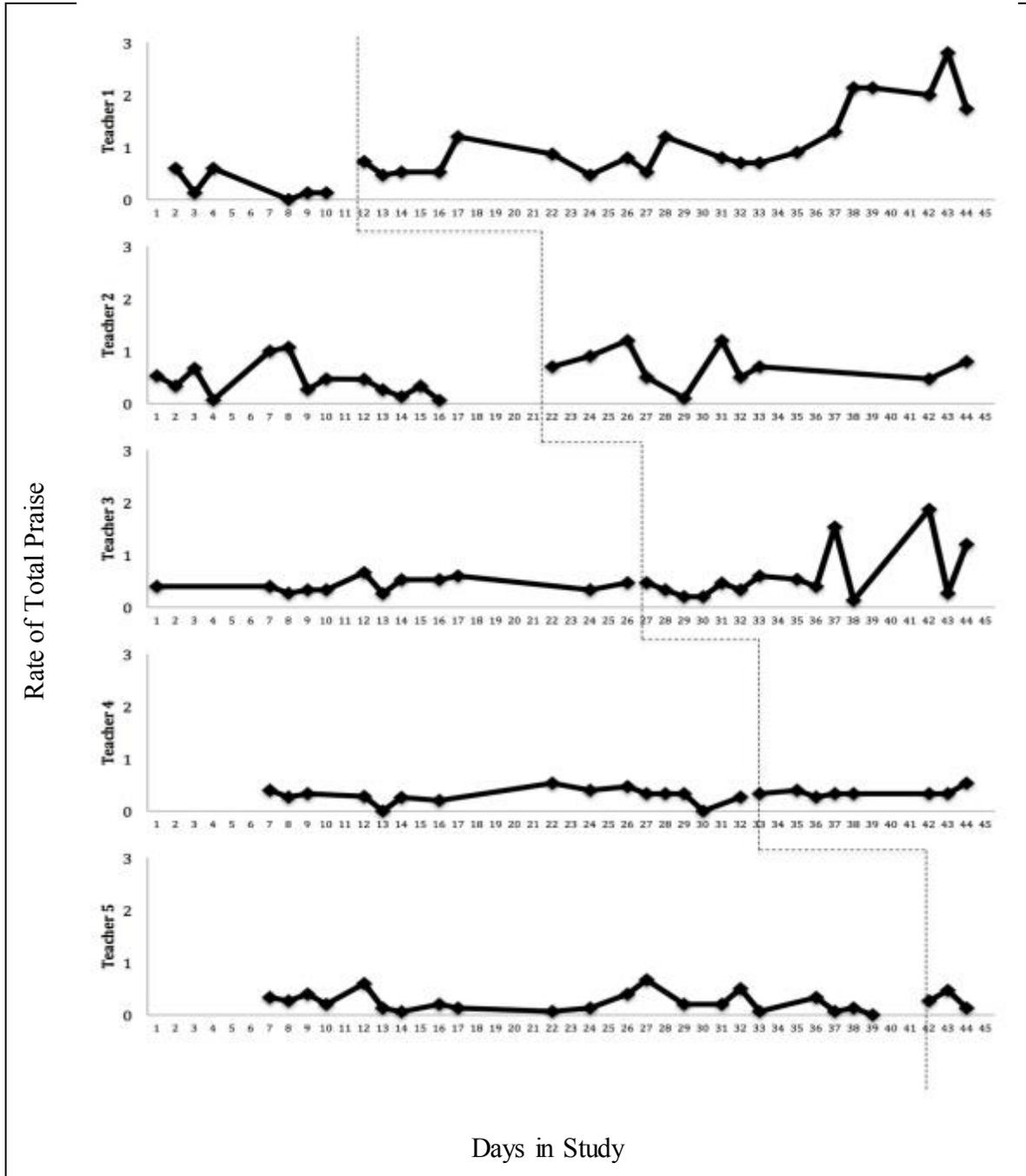


Figure 2. BCIO-R: Overall Rate of Praise

Table 4.

Slope Across Phases for BCIO-R: Overall Rate of Praise

	Baseline Phase Slope	Intervention Phase Slope	Baseline Phase Mean	Intervention Phase Mean
Subject 1	-.04	.03	.27	1.13
Subject 2	-.02	.005	.43	.71
Subject 3	.002	.04	.43	.61
Subject 4	.001	.006	.29	.36
Subject 5	.002	-.06	.24	.29

Table 5.

Advanced Empirical Analyses of Intervention Effects on Rate of Praise

	PND	PEM	Tau U	PAND	Cohen's d
Teacher 1	43%	43%	83%	-	-
Teacher 2	20%	90%	51%	-	-
Teacher 3	21%	50%	0%	-	-
Teacher 4	0%	0%	28%	-	-
Teacher 5	0%	67%	27%	-	-
All Subjects	-	-	-	36%	-0.6

Note. PND & PEM <50% indicates no intervention effect. Tau-u <65% indicates weak or small effect. Cohen's *d* of .2 = small, .5 = medium, & .8 = large.

Table 6.

BCIO-R: Overall Rate of Praise Contingency Table Analysis

	Intervention	Baseline	Total
Higher	13.26%	31.82%	45.08%
Lower	31.82%	23.10%	54.92%
Total	45.08%	54.92%	100.00%

Table 7.

User Rating Profile-Intervention Mean Item Score

	Accept	Understand	Feasible	System Support	Total
Rater 1	5	5	5	4	5
Rater 2	5	4	4	4	4
Rater 3	3	5	4	3	4
Rater 4	3	3	3	4	3
Rater 5	5	6	5	4	5
ALL	4	5	4	3	4

Note. Ratings were anonymous. Therefore, Rater number does not match intervention phase teacher/subject number. Interpretation guidelines for The URP- I states “Higher scores are more desirable” (1 = Strongly Disagree, 2 = Disagree, 3 = Slightly Disagree, 4 = Slightly Agree, 5 = Agree, and 6 = Strongly Agree)

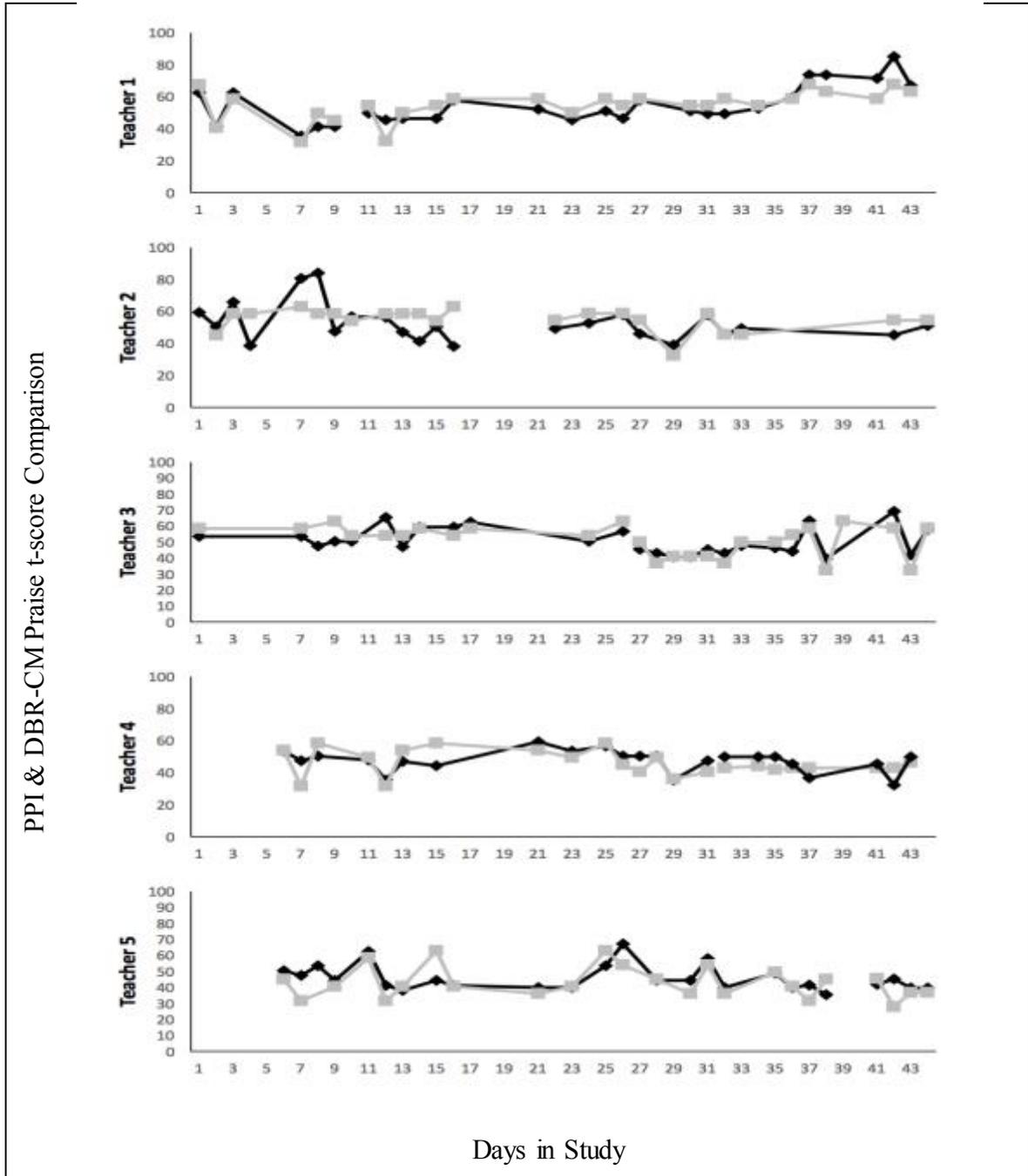


Figure 3. Visual Presentation of BCIO-R: Overall Rate of Praise and DBR-CM Praise t-scores. Note. DBR – CM t-scores presented in grey.

VITA

Wesley A. Sims is a Nationally Certified School Psychologist that has worked in a variety of schools serving diverse populations since 2005. Mr. Sims was born in Kansas and has lived in eight states. He attended high school and college in Texas, earning a Bachelor of Arts from the University of North Texas. Mr. Sims attended graduate school in Boston at that University of Massachusetts Boston. After earning a Master of Education and Certificate of Advanced Graduate Study, Mr. Sims began his career as a practicing school psychologist in schools in Columbia, MO. As a practicing School Psychologist, Mr. Sims garnered extensive experience facilitating response to intervention, problem-solving, and data teams. This experience led Mr. Sims to develop and implement a problem-solving process that has been implemented successfully in several Missouri schools and presented locally, regionally, and internationally. In addition to his experience with systems-level change, problem-solving teams, and RTI, Mr. Sims has significant experience with school-based consultation, academic and behavior assessment and intervention, eligibility determinations, and oversight of special education processes, and timelines. Mr. Sims has expanded his professional work to research in the areas of behavior assessment, classroom management assessment, and implementation science (developing sustainable evidence-based practices, validating practitioner practices, and generally bridging the gap between research and practice).

In his most recent placement as a School Psychologist, Mr. Sims introduced school psychology to a rural Missouri school district that had previously never had a school psychologist. For these efforts, in part, Mr. Sims was recognized as the School Psychologist of the Year for 2013 – 2014 by the Missouri Association of School

Psychologists (MASP) and a nomination for the National Association of School Psychologists' (NASP) School Psychologist of the Year award. Mr. Sims has held several leadership positions within his state's school psychology association, including President, President-Elect, and Conference Chair. Mr. Sims continues to consult with schools around improving individual, group, and systems level practices as a technical assistance coach for the National Center for Intensive Intervention.

His time at the University of Missouri in the completion of his doctoral training was highly influential in the development of an independent research program. His professional goals have expanded to include conducting high-quality applied research that will impact the daily practices of education and psycho-educational service provision. His experience working on multiple large-scale, federally funded research projects served as exemplary models for developing and executing a impactful research program. Generally, Mr. Sims' research focuses on assessment and intervention of educational service implementers and the processes that guide service delivery. Inefficient or inconsistent use of procedures or mechanisms that guide service delivery render powerless the wealth of evidence-based interventions available to educators. His work focuses on the assessment and promotion of practices and mechanisms that promote efficient, effective, and consistent service delivery within multi-tiered systems of support. This translates explicitly to my work developing, evaluating, and using the Direct Behavior Rating – Classroom Management and evaluating school-based problem-solving teams (SB PST) while refining and evaluating an alternative SB PST model. This research is grounded in a problem-solving orientation, which aligns with a prevention and early intervention perspective. The goal of this research is developing practical,

productive, and durable changes that promote efficient and effective general and targeted service delivery. I believe my research agenda is particularly well suited to the training emphasis of your programs as it is conceptualized through problem-solving and data-based decision making frameworks, which protect and promote ecological and contextual factors that influence student functioning.