UNIVERSITY OF CALIFORNIA
RIVERSIDE

Evaluating the Effects of Performance Feedback-Based Professional Development on Educator Classroom Management Practices

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

Doctor of Philosophy

in

Education

by

Theresa Marie Canché

March 2024

Dissertation Committee:
Dr. Wesley Sims, Chairperson
Dr. Katherine Meltzoff
Dr. Soojin Park

The Dissertation of Theresa Marie Canché is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ABSTRACT OF THE DISSERTATION

Evaluating the Effects of Performance Feedback-Based Professional Development on Educator Classroom Management Practices

by

Theresa Marie Canché

Doctor of Philosophy, Graduate Program in Education
University of California, Riverside, March 2024
Dr. Wesley Sims, Chairperson

Effective classroom management (CM) is critical to promoting student success, as there is a clear connection between teacher CM practices and student academic, social, and behavioral outcomes. Research demonstrates that children spend more time academically engaged, progress more rapidly, and have higher levels of academic achievement when they are in a well-managed classroom. Despite the importance of CM and its influence on student outcomes, pre- and in-service training in this area appears lacking. To address this issue, some scholars have proposed the use of a data-driven, multi-tiered approach to support educational service delivery, including professional development (PD) activities designed to promote use of effective CM practices. However, in current, applied practices, PD opportunities fall largely at two ends of the spectrum of time, energy, and resources required. The goal of the present study was to explore the utility of a moderately intensive, Tier II level, PD approach within a

multitiered system of educator support (MTSES) framework. A combined multiple baseline, ABC single case design across participants was used to evaluate the effects of two, PD approaches, performance feedback (PF) and performance feedback with self-monitoring (PF+SM), on the CM practices of four volunteer participants. Classroom management was measured through daily 20-minute observations using the Direct Behavior Rating- Classroom Management (DBR-CM). During the first intervention phase, participants received brief verbal and visual performance feedback before each observation period and during the second intervention phase, a self-monitoring component was added to these procedures. Results of visual and empirical analyses indicated that the PF only intervention was effective in increasing teachers use of evidence-based CM practices as measured by DBR-CM total score. Slight improvements in CM practices from the first intervention to the second intervention (PF + SM) were observed for individual participants but no effect of this intervention was demonstrated. Several important limitations and suggestions for future research are discussed.

Introduction **1**

Classroom Management 3

Evidence-based CM Practices *4*

CM and Student-level Outcomes *7*

CM Training 9

Preservice CM Training *9*

Inservice CM PD *11*

Shortcomings in Typical Educational PD *12*

Consultation-based PD 14

School-based Consultation *14*

Coaching *16*

Performance Feedback *18*

Consultation-based PD Effectiveness *20*

Coaching and PF Implementation Considerations *23*

Self-monitoring 26

SM Effectiveness *26*

SM Implementation Considerations *28*

Multi-Tiered System of Educator Support 29

Tiered CM PD Activities *30*

Expanding CM PD Options *32*

Importance of Data in MTSES 33

CM Assessment 34

DBR-CM Background 36

Development *36*

Validation *37*

DBR-CM Supported PD 38

Prior DBR-CM PD Work *39*

Study Purpose 41

Methods 42

Participants 42

Measures 42

Demographic Survey *42*

Direct Behavior Rating –CM *43*

Modified User Rating Profile-Intervention Revised *44*

Fidelity Measure *45*

Design 45

Dependent Variable *48*

Procedure 49

Data Collection *49*

PD Activities (i.e., treatment/intervention) 54

Results 61

Participant Demographics 61

Fidelity 61

Visual Analysis 62

Teacher 1 ..... *66*

Teacher 2 ..... *67*

Teacher 3 ..... *68*

Teacher 4 ..... *69*

Empirical analyses 69

PND *70*

PAND *70*

Tau-U *71*

Interobserver Agreement 72

Social Validity 73

Discussion 76

Research Hypotheses 76

Performance Feedback Intervention *76*

Performance Feedback & Self-Monitoring Intervention *79*

Acceptability: Performance feedback *83*

Acceptability: Performance Feedback & Self-Monitoring *84*

Limitations 85

Future directions 88

Conclusion 89

References 93

Appendix B: DBR-CM SR 100

Appendix C: DBR-CM ER 101

Appendix D (DBR-CM ER Example PF Graph: Phase B and Phase C) 102

Figure 1 66

Table 1..... 64
Table 2..... 72
Table 3..... 73
Table 4..... 75
Table 5..... 75

Introduction

Classroom Management (CM) may be the most significant factor in student success and teacher retention. Wang and colleagues (1993) rated CM to be first in terms of impact on student achievement as there is a clear connection between teacher CM practices and student academic, social, and behavioral outcomes. In particular, problematic student behavior generally predicts low academic achievement, school dropout, and drug abuse in adolescence and is associated with learning problems (Walker et al., 2000). Furthermore, research has demonstrated that poorly managed classrooms have higher levels of off-task behavior and student disruptions which cause a loss of instructional time (Horne, 1980; Emmer & Stough, 2001). In contrast, Oliver & Reschly (2007) note effectively managed classrooms experience more positive outcomes and teachers spend more spend more time teaching and less time managing student behaviors and disruptions. Despite the importance of CM and the outcomes associated with it, pre- and in-service training in this area appears lacking, which can leave teachers feeling inadequately prepared and lead to high rates of burnout (Bayar, 2014; Brouwers & Tomic, 2000; Corcoran, 1995; Ingersoll & Smith, 2003). When surveyed, teachers persistently note deficiencies in their pre-service CM training (Christofferson & Sullivan, 2015; Cooper et al., 2018; Flower et al., 2017; Freeman et al., 2014; Poznanski et al., 2018). To address this issue, some scholars have proposed the use of a data-driven, multi-tiered approach to support educational service delivery, including professional development (PD) activities designed to promote the use of effective CM practices (Grasley-Boy et al., 2021; Simonsen, et al., 2014; Sims et al., 2021). Like those

supporting students directly, a multi-tiered system of educator support (e.g., MTSES; Sims et al., 2021) would provide varying levels of support for a wide range of specific CM skills (Sims et al., 2021; Grasley-Boy et al., 2019). However, in current, applied practices, PD opportunities seem to fall at two ends of the spectrum of intensity or the amount of time, energy, and resources required for implementation. Found at one end are brief, infrequent, lecture-style in-service workshops, which are relatively easily delivered, but do not meet the needs of all teachers and are notoriously ineffective. At the other extreme of the spectrum are more effective, but also more and resource intensive activities such as consultation-based coaching and performance feedback (PF) PD activities. While notably less time and resource intensive, in-service, lecture-style trainings focus on information dissemination, often leaving teachers in need of some follow-up support to successfully implement skills and interventions they learn about in these PD workshops and training. Unfortunately, most coaching- and performance feedback-based PD activities may not be feasible in all schools. Because there are limited options for PD opportunities that fall between these two extremes of too intensive and not intensive enough, further research on the subject is warranted. Sims (2016) proposed a less time and resource intensive performance feedback intervention to improve teachers' CM practices and results were mixed. The present study aims to modify this less intensive performance feedback intervention and add a self-monitoring (SM) component to improve teachers' CM practices through a feasible and effective intervention. Additional PD options of moderate intensity or invasiveness, consistent with Tier II level support appears needed. Such PD activities would serve to further bridge the gap between

the support levels currently available for teachers who need support but in a more time and resource frugal format. The purpose of the present study is to evaluate the effects of CM focused PF and SM PD approaches using the DBR-CM. Both PF and SM represent less intensive, Tier II PD approaches to improving educator use of evidence-based CM practices.

Classroom Management

CM emerged as a sub-field within educational research in the 1950s, which has resulted in the accumulation of decades worth of literature on the topic (Martin et al., 2016). Early research on CM was divided between a behavioral perspective or an ecological perspective (Martin et al., 2016). The behavioral theory of CM focused on the individual, observable actions of the teacher that either reinforced or punished student behavior, while the ecological theory analyzed the classroom environment as a whole (Martin et al., 2016). Current conceptualizations of CM are more closely aligned with the ecological perspective, although definitions and theoretical approaches remain varied (Martin, et al., 2016).

Jones (1996) emphasized the comprehensiveness of CM by broadly defining it as establishing and maintaining order, designing effective instruction, managing students as a group, responding to individual needs, and effectively handling discipline and adjustment of each student (Emmer & Stough, 2001). Similarly, Back and colleagues (2016) define CM as the practices employed by teachers to create and maintain supportive and productive classroom environments. Everson and Weinstein (2006) also define CM broadly as teacher actions that create a supportive classroom environment but

add that these actions should facilitate academic and social-emotional learning. Although these definitions are varied, CM consistently includes teacher actions that maintain a supportive environment, establish order, and promote learning and cooperation (Emmer & Stough, 2001).

Evidence-based CM Practices

There are a variety of evidence-based (EB) CM practices that have been identified in the literature (Freeman et al., 2014). For example, explicitly teaching classroom routines and expectations, providing prompts and precorrections for appropriate behaviors, delivering behavior-specific praise, high rates of opportunities to respond, and maintaining a greater number of positive to negative interactions with students are all empirically supported practices that improve student outcomes (Simonsen et al., 2020). There are many approaches to classroom management and each one differs in how behavior should be conceptualized and how to approach teaching and modifying student behaviors (Freeman et al., 2014). Some evidence-based perspectives on CM emphasize teachers' self-regulation/control (Emmer & Stough, 2001), others focus on creating a sense of membership of the classroom community (Schafer, 2006), while others stress the importance of constructing an image of caring and authority (Wubbles, et al., 2006). However, taken individually, these perspectives only address individual components of CM and ignore the broader context of CM as described above. These individual practices, which are often associated with a specific theory (e.g., behaviorism, social learning), should be combined and flexibly utilized in specific contexts to create successful classroom environments.

In a review of the literature on effective classroom rules, Alter & Haydon (2017) found that classroom rules are a simple yet effective evidence-based way to prevent challenging behaviors before they occur. The two most important characteristics of this empirically supported classroom management practice are teaching the rules to students and tying rules to positive or negative consequences (Alter & Haydon, 2017). Behavior-specific praise (BSP) is another evidence-based classroom management practice that prevents and decreases student problem behavior (Floress et al., 2017, Royer et al., 2019). Decades of research have proven the many positive effects of teacher delivered BSP including enhanced student-teacher relationships, positive teaching environments, increased instructional time, and positive student outcomes (Becker et al., 1967; Conroy et al., 2009; Reinke et al., 2008; Sutherland et al., 2000; Sutherland et al., 2008). Increasing teacher directed opportunities to respond (TD-OTR) is a classroom management practice that elicits student responses and is followed by teacher feedback (MacSuga-Gauge & Simonsen, 2015). MacSuga-Gauge & Simonsen (2015) reviewed 15 empirical studies and found that increasing TD-OTR supported positive academic and behavioral student outcomes. One of the simplest, evidence-based classroom management practices is precorrection which is used to increase appropriate and prevent inappropriate classroom behavior (Faul et al., 2012). Precorrections are delivered before desired behavior is expected often through verbal prompts, reminders, or modeling (Faul et al., 2012). In research, precorrections have demonstrated an increase in appropriate behavior, a decrease in inappropriate behavior, and have proven effective for a variety of ages (i.e., preschool to adult) and ability levels (those with and without disabilities) (Faul

et al., 2012; Flood et al. 2002; Wilder & Atwell, 2006). Other classroom management practices that are empirically supported include reprimands, appropriate instructional pace, varied instructional methodologies, increasing student engagement, and using attention signals (Freeman et al., 2014; Pianta & Hamre, 2009; Reinke et al., 2013; Simonsen et al., 2010; Solomon et al., 1996).

Simonsen and colleagues (2008) reviewed the extensive literature on evidence-based CM practices and suggested five categories: (a) maximizing structure, (b) establishing and teaching positive expectations, (c) actively engage students in instruction, (d) implement a continuum of strategies to increase appropriate behavior, and (e) implement a continuum of strategies designed to decrease inappropriate behaviors (Simonsen et al., 2008). These five critical features of evidence-based CM practices are meant to be combined to create and maintain a supportive, productive learning environment for students. More broadly, positive preventative CM practices like clear communication of behavioral expectations and behavior-specific praise are strongly supported by evidence, while more negative reactive practices like punishing inappropriate behaviors and excluding students from instruction are not effective or evidence-based and often exacerbate the problem (Kennedy & Jolivette, 2008; Sutherland et al., 2000).

The research literature on CM also revealed a shared theme not captured in any one, easily observable behavior related to the quality of student-teacher relationships, consistent with the concept of rapport. In this context, rapport could refer to the quality of the student-teacher relationship and emphasizes mutual trust, acceptance, and emotional

affinity (Sims et al, 2021). A mutually positive, respectful, and warm relationship between the teacher and students is a key element in creating a successful classroom, and the importance of teachers making a conscious effort to create an atmosphere of warmth and acceptance is repeatedly highlighted in the literature (Bracken & Fischel, 2006; Pianta & Hamre, 2009; Sprick et al., 1998). Pianta & Hamre (2009) discuss the importance of positive classroom climate, teacher sensitivity, and regard for student perspectives in the development of a child's social and emotional functioning. High quality student-teacher relationships, or rapport, is a positive predictor of behavioral and academic development by fostering motivation and connection to others (Hamre & Pianta, 2001). Rapport is therefore an essential element of effective classroom management.

CM and Student-level Outcomes

There is a clear connection between teacher CM behavior and student outcomes in classrooms. Effectively managed classrooms experience more positive outcomes and teachers spend more time teaching and less time managing student behaviors and disruptions (Oliver & Reschly, 2007). Children spend more time academically engaged, progress more rapidly, and have higher levels of academic achievement when they are in a well-managed classroom versus a poorly managed one (Brophy, 1985). Other characteristics of a well-managed classroom include lower 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 (La paro & Pianta, 2003).

Research has demonstrated that poorly managed classrooms have higher levels of off-task behavior and student disruptions which cause a loss of instructional time. The negative effects of a poorly managed classroom are both immediate and long-term. Classrooms that are not well-managed are associated with long-term negative academic, behavioral, and social outcomes (Horne, 1980; Kellam et al., 1998). Students in classrooms that are not managed effectively do not learn positive social behaviors, and the important role that rules and consequences play both within and beyond the classroom setting (Horne, 1980).

Importantly, previous research has also established a clear link between behavior and academic achievement. Behavior problems generally predict low academic achievement, school dropout, and drug abuse in adolescence and are associated with learning problems (Walker et al., 2000). Students with behavioral problems and poor school readiness skills are at greater risk of school failure, delinquency, and other risky adolescent behaviors like heavy drinking, risky sexual practices, and violent delinquent acts (Walker et al., 2000). The association between CM and classroom behavior coupled with the correlation between classroom behavior and academic achievement present a convincing argument for addressing teacher CM behavior. In seeking out ways to improve CM practices, the goal is to mitigate some of these negative outcomes associated with poor CM.

Despite the extensive literature supporting the use of evidence-based CM practices, research suggests implementation of EB CM practices in classrooms may be infrequent (Oliver et al., 2015). Teachers often rely on their own experiences to inform their CM strategies, resulting in practices that lack a strong evidence-base (Brophy & McCauslin, 1992). The deficient use of evidence-based CM strategies being utilized in schools can be largely attributed to the lack of training teachers receive in their programs (Oliver & Reschly, 2010).

CM Training

Preservice CM Training

Like most professions, development of teaching skills and competencies, including those in CM, occur primarily in a secondary educational setting (e.g., college, university). Unfortunately, although there is a large body of research regarding the importance of effective, evidence-based CM practices, teacher preparation programs appear to spend relatively little instructional time on these skills (Oliver & Reschly, 2010). Both general and special education teachers have consistently and persistently reported non-pedagogical CM practices (e.g., preventing and managing problem behavior) as the area of greatest deficit within their pre-service training (Begeny & Martens; Freeman et al., 2014; Pindiprolu et al., 2007; Stough, 2006). While most states have broad policies that require teacher preparation programs to have instruction in CM, only about half require instruction in research-based practices, and the extent of the instruction is not specified (Freeman et al., 2014). For example, Merrett and Wheldall (1993) found only 18 percent of teachers report learning CM skills as part of their teacher

preparation program. While somewhat more favorable, Freeman and colleagues (2014) found that less than half of general education teacher preparation programs included any instruction in evidence-based CM practices. Of the programs that included some type of instruction in CM, many focused only on reactive practice, or how to respond to problem behaviors rather than teaching proactive strategies meant to prevent these behaviors from happening (Oliver & Reschly, 2010). This limited perspective is problematic in that it does not align with contemporary views of CM (i.e., eclectic, multidimensional conceptualization) which recognizes and emphasizes the importance of development and use of preventative CM strategies in addition to reactive ones (Brophy, 1985; Kennedy & Jolivette, 2008; Sutherland et al., 2000). Additionally, of the programs that did include explicit instruction on CM practices, this content appeared to be covered as a small component within one or two classes, rather than in a more substantial, comprehensive format such as a dedicated course (Freeman, 2014). To this point, Oliver & Reschly (2010) found that only 27 percent of university special education programs in a midwestern state taught CM as an independent course.

A critical review of preservice teacher CM training practices suggests that there is ample room for improvement, a conclusion that is problematic for numerous reasons. First, deficiencies in teacher preparation likely contributes to attrition from the profession (Freeman et al., 2014). Next, such shortcoming in formal preservice training results in teachers seeking out alternate sources to support development of their CM skills. For example, classroom teachers identified mentoring and fieldwork as the most common sources of CM training (Christofferson & Sullivan, 2015). Given mentoring teachers

likely experienced poor preservice CM training and consistent with literature indicating widespread use of non-evidence-based CM practices used by teachers, this form of PD appears less than desirable. This in turn leads to the numerous poor outcomes noted for students in poorly managed classrooms. Lastly, collectively this forces districts, buildings, and administrators to supplement preservice teacher training with in-service PD activities to improve teacher CM practices as well as dedicate infinite educational resources to address the student-level challenges associated with poor CM practices (e.g., problematic behavior, poor academic achievement, poorer social-emotional functioning; Horne, 1980; Kellam et al., 1998; Oliver & Reschly, 2007; Walker et al., 2000).

Inservice CM PD

Despite the importance of CM and the outcomes associated with it, teacher preparation programs place little emphasis on this aspect of training which can leave teachers feeling inadequately prepared to face these challenges in the classroom (Bayar, 2014; Corcoran, 1995). Unfortunately, this leaves the challenge of identifying effective, relevant, and feasible ways to improve teacher behavior to administrators (Myers et al., 2011). Broadly, PD is the additional education that teachers receive after completing their certification and while they are employed (Stough & Montague, 2015). PD can fill in gaps of knowledge, keep teachers updated on current theory and legal mandates, and allow for reflection of teachers own educational practices and beliefs (Stough & Montague, 2015). Like student instruction, PD can be conceptualized based on the desired outcome for participants relative to the topic, practice, or skill targeted. PD activities focused on the *Acquisition* of knowledge or skills would focus on dissemination

of information (Freeman et al., 2017). In contrast, PD activities seeking to promote knowledge or skill *Fluency,* would emphasize its repeated, efficient practical application or use (Freeman et al., 2017). Similarly, *Generalization*-oriented PD activities would seek to support use of acquired and fluent knowledge and skills in novel settings or situations (Freeman et al., 2017). While a variety of PD activities across each of these outcomes (i.e., *Acquisition, Fluency, Generalization)* exist, districts tend to rely predominantly on *Acquisition*-oriented PD activities. Educational PD often takes the form of one-time workshops, conferences, or dyadic presentation-style training sessions (Dunst et al., 2015). In-service trainings and workshops are often isolated events that are not tied to ongoing PD opportunities (Parise & Spillane; 2010). These one- or two-time events are typically informational sessions that focus on disseminating materials rather than building skill fluency (Sugai & Horner, 2009). Teachers participate in these training sessions through their school district, through universities, conferences, or regional workshops (Dunst et al., 2015). Unfortunately, these informational, in-service professional development activities that districts and administrators typically employ, appear to have little tangible impact on educator practices (Birman et al., 2007; Guskey, 2000; Guskey & Yoon, 2009).

Shortcomings in Typical Educational PD

Literature suggests typical approaches to PD in educational settings take the form of in-service training or informal mentorship by other teachers with more teaching experience (Bayar, 2014; Corcoran, 1995; Parise & Spillane, 2010). For example, there are a variety of PD programs (e.g., Classroom Organization and Management Program,

Teacher Effectiveness Training) that teachers are exposed to, but there is often little to no follow up or continuity with the programs making it difficult for teachers to realize the benefit from the practices discussed in such trainings (Dunst et al., 2015). In-service trainings, or 'sit and get' or 'one shot' workshops, are often not effective for most teachers (Hunzicker, 2011). Teachers are unlikely to remember most of the information gained and are even less likely to apply it once they are back in the classroom (Hunzicker, 2011). Training and PD opportunities that aim to improve teacher's CM behavior are not excluded from these findings, as they rarely result in lasting change in teacher practices (Corcoran, 1995; Elmore, 2005). This ineffectiveness may be attributed to multiple considerations within this approach. First, as noted previously, this information dissemination approach is not genuinely designed to provoke changes in teacher practices (i.e., *Fluency*, *Generalization*), typically it seeks to facilitate exposure to new information (i.e., *Acquisition*). Next, typical educational PD activities are provided en masse and do not offer specific, targeted support depending on the teachers' skills and level of need (Corcoran, 1995; Sugai & Horner, 2009). The informational, one-size-fits-all approach does not account for varied strengths, weaknesses, background, experiences, wants, or needs as part of the PD provided. Lastly, this approach lacks follow up and specific targeted support following presentation of PD information (Sugai & Horner, 2009). More specifically, in-service workshops typically fail to include any mechanisms by which to assess implementation integrity and on-site support for implementation of the learned skills.

In summary, many teachers do not receive adequate training in CM prior to beginning their teaching careers and often feel unprepared to manage student behaviors in their classrooms (Freeman et al., 2014; Reinke et al., 2008; Wagner et al., 2006). This inadequate training can lead to increased risk of teacher burn-out and attrition from the profession (Brouwers & Tomic, 2000; Ingersoll & Smith, 2003). CM practices impact student behavior which impacts academic achievement and other long term student outcomes. Therefore, it is important for future research to seek out ways to improve CM practices. While traditionally adopted PD approaches appear largely ineffective beyond dissemination of information, PD activities that emphasize data-driven ongoing feedback and targeted support offer educators much more desirable, durable outcomes (Pindiprolu et al., 2007).

Consultation-based PD

Pre-service teacher training and typical in-service approaches to professional development do not adequately prepare teachers to effectively manage their classrooms (Freeman et al., 2014; Wagner et al., 2006). However, consultation-based PD activities such as coaching and performance feedback are more effective methods of supporting teachers use of empirically supported classroom management practices in the classroom. These consultation-based PD activities move beyond acquisition of knowledge and promote fluency and generalization of these learned skills.

School-based Consultation

School-based consultation (SBC) is a widely used method to address student academic and behavioral challenges that arise in the classroom (Bergan, 1977). There is

also a strong evidence base supporting the effectiveness of consultation in schools for improving teacher and student outcomes compared with didactic training (Kratochwill et al., 2014; Kratochwill & Bergan, 1990). Through SBC school psychologists and other professionals can work with teachers and other adults to indirectly address CM concerns which is often the best way to reach the greatest number of students (Newman & Morrison, 2019). Consultation is an efficient way to help teachers and other professionals develop the knowledge, skills, confidence, and objectivity they need to be effective (Newman & Morrison, 2019). Consultation allows consultants to give away their expertise to teachers who can then use that knowledge to reach their current and future students. By providing consultation to teachers, classroom management experts can reach a much larger number of students than they would by intervening on individual students' behavior.

Consultation is also less time and resource intensive because the consultee (i.e., the teacher) is typically the one responsible for data collection and implementation of the intervention or strategy (Erchul, 2019). Another benefit of this PD activity is that it can be integrated within multi-tiered systems of support (MTSS) to incorporate preventative and early interventions (Erchul, 2019). Consultations can be delivered to individual teachers and teams at each level of MTSS to promote universal screening, data-based decision making, progress monitoring, and evidence-based academic and behavioral interventions (Newman & Morrison, 2019). However, some models of consultation place an emphasis on the students and attribute difficulties to within a student rather than focusing on the teacher and the instructional environment (Reinke et al., 2008). Targeting

the entire classroom system to promote the use of effective classroom management practices delivered to all students is more efficient than targeting individual students (Reinke et al., 2008). Increased use of effective classroom management practices is likely to reduce current student behavioral and academic difficulties and prevent future student problems on a broader scale (Reinke et al., 2008). For these reasons, consultation can be a cost effective and useful tool for improving the classroom as well as individual student outcomes.

A consistent limitation of most approaches to school-based consultation (SBC) is that individual students' problematic behaviors are the focus of change rather than the educator and the classroom system (Reinke et al., 2008; Sheridan et al., 1996). This focus on students can pathologize difference and attribute problematic behavior to within student factors (Sheridan et al., 1996). Making students the focus of change in consultation results in interventions that target individual students rather than classroom management practices as a whole. However, when the focus of the consultation relationship is on implementing interventions that improve educators and the classroom environment, there is potential to dramatically impact a large number of students (Reinke et al., 2013).

Coaching

In contrast to the oft employed in-service workshop approach to PD, coaching offers a more actively engaging, long-term, job-embedded approach to PD (Denton & Hasbrouck, 2009; Desimone & Pak, 2017). Coaching is used in schools for a variety of reasons including new teacher training, ongoing teacher learning, and helping teachers

adapt instruction to new standards (Desimone & Pak, 2017). While consultant-based coaching (i.e., not teacher mentor driven) to improve CM practices is relatively new, instructional coaching (IC) offers some insight as to what coaching with teachers might look like. Instructional coaching in school settings offers a way to support teachers in their efforts to provide high quality teaching (Denton & Hasbrouck, 2009). Coaching is a supportive PD practice where someone with specialized knowledge in a specific area works directly with a classroom teacher to change current practices and improve skills (Denton & Hasbrouck, 2009). Coaches engage in a variety of activities in schools including observing lessons and providing feedback, modeling effective teaching strategies, advising and supporting the improvement of lesson planning and implementation, assisting with managing the classroom, and developing and monitoring goals and plans for improvement (Denton & Hasbrouck, 2009).

Instructional coaching is not limited to addressing academic instructional practices. Instructional coaching can also support teachers in improving their classroom management practices (Devine et al., 2013; Knight, 2018; Reinke et al., 2009). As with academically focused IC, classroom management IC requires coaches to be knowledgeable about a particular subject area. For IC that targets classroom management, coaches must be familiar with research-based interventions and strategies of classroom management (Reinke et al., 2009). The role of the coach is to recognize areas that can be targeted to produce positive change, identify interventions, demonstrate the strategy (but not by modeling in a teachers classroom with children present) observe the teacher, analyze data, and review results (Reinke et al., 2009; Knight, 2018). Coaches

are most effective when they can establish good rapport and open communication and provide help and support to educators while they learn to fluently implement the strategies they have learned (Knight, 2018; Reinke et al., 2009).

Performance Feedback

Like coaching, PF, or on-going, data-based feedback on current performance of targeted behaviors is one of the best ways to improve implementation of evidence-based practices (Reinke et al., 2007). PF was defined by Kluger & DeNisi (1996) as "actions taken by external agents to provide information regarding some aspects of one's task performance" (p.255). However, in the specific context of schools, performance feedback is defined as "monitoring a behavior that is the focus of concern and providing feedback to the individual regarding that behavior" (Noell et al., 2005). The purpose of performance feedback is to improve the target behaviors through the collection and presentation of data to the individual (Cavanaugh, 2013). According to the literature, performance feedback is most effective when it focuses on tasks rather than personal qualities, and targets behaviors that the individual being observed views as feasible, necessary, and based on appropriate goals (Cavanaugh, 2013). In research, performance feedback is usually implemented by researchers or university students with extensive training and/or experience (Cavanaugh, 2013). However, in practice, performance feedback is often implemented by available staff members like administrators or school-based coaches who do not always have the same training, availability, and/or expertise (Cavanaugh, 2013). This discrepancy in implementation makes it difficult to determine whether the positive results found in research would generalize to an actual school setting

(Cavanaugh, 2013). Furthermore, performance feedback in schools often takes the form of administrator evaluations which occur infrequently, are highly subjective, and collect data using instruments that have no normative information or standardization (Sims, 2016). This type of performance feedback does not provide meaningful data that can be used to monitor and change teacher behavior in the classroom (Sims, 2016).

Specific iterations of PF can take a variety of forms (e.g., visual, written, or verbal feedback) with varying levels of intensity. There is a large body of literature to support the effectiveness of PF in a variety of formats and intensities. On the less intensive end of the PF spectrum, Sutherland and colleagues (2000) found that teachers were able to increase their rates of behavior specific praise (BSP) when provided with only verbal feedback about their rate of BSP. Teachers were reminded of their goal prior to each observation and then immediately following the observation they were provided with verbal feedback regarding their use of BSP. In another study by Duchaine and colleagues (2011), teachers received verbal performance feedback on their rate of BSP after every third observation session and written performance feedback after every session. As a result of this verbal and written performance feedback, teachers' rates of BSP increased (Duchaine, Jolivette, & Frederick et al., 2011). Reinke and colleagues (2007) discovered that visual performance feedback (VPF) alone increased teachers' use of behavior-specific praise for both target students and classroom peers. In this study, teachers were provided with a graph at the beginning of each day which displayed their use of behavior-specific praise. This VPF was not paired with any verbal feedback or interaction with the

researchers. Instead, group consultation meetings focused on BSP were held with all teachers at three different points in the study (Reinke et al., 2007).

More intensive forms of performance feedback have also proven to be successful at changing teacher behavior. In one study by Thompson and colleagues (2012), video self-monitoring was used to successfully improve teachers' rates of BSP. In this study, teachers recorded a brief (15-25 minute) video segment of themselves teaching and then watched it back while counting the total number of BSP's they used. Pisacreta and colleagues (2011) used modeling and performance feedback to train teachers to implement a 1:1 ratio of praise-to-behavior correction during instructional time. The experimenter modeled how and when they wanted the teacher to praise students' appropriate behaviors. After modeling, teachers were observed during implementation and then received brief, verbal, performance feedback after each observation session. Results indicated that after training, teachers who initially gave little verbal praise were able to maintain a 1:1 praise-to-behavior correction ratio throughout the intervention. Noell and colleagues (2005) found that after an initial consultation with teachers, performance feedback led to better implementation of treatment plans and student outcomes. In this study, PF was provided daily at first, then faded to every other day, and finally decreased to weekly PF. These findings demonstrate that a wide range of performance feedback methods can be effective in changing teacher behavior.

Consultation-based PD Effectiveness

There is a plethora of research demonstrating that teachers benefit from the additional support of a coach when implementing new strategies and instructional

practices. Coaching is a more targeted and engaging PD practice than traditional workshop trainings that often lack individualization and continuity. Coaching offers teachers targeted and ongoing support in areas of weakness. Coaches can provide specific feedback regarding teachers' instructional practices and implementation of new skills and strategies which improves teachers' belief in their own abilities and implementation fidelity (Reinke et al., 2014). Research on coaching dating back to the 1980s reported positive outcomes in teachers lesson planning and organization, CM strategies, and meeting student instructional goals (Desimone & Pak, 2017). Kraft et al. (2018) conducted a meta-analysis of the evidence for teacher coaching on teacher instructional practices and student academic achievement. There were 60 studies included in the analysis and the results indicated that coaching improved both teacher instruction and student academic achievement compared with the typical improvement seen across the first 5-10 years of teaching (Kraft et al. 2018). Research has also demonstrated that when teachers are provided coaching to support their implementation of behavioral interventions, they are more likely to implement the intervention effectively, maintain the skills they learned, and report higher self-efficacy (Reinke et al., 2014). In a review of the literature on the use of coaching to support implementation of social behavioral interventions, Stormont and colleagues (2015) found that 86 percent of studies reported positive findings. Joyce and Showers (1982) examined the use of coaching in supporting teachers in implementing new skills and approaches to teaching and found that coaching enabled most teachers to use a new teaching model/approach fluently and appropriately. A majority of the studies that examined the use of coaching to improve aspects of

teachers' CM specifically focused on increasing teachers' use of praise. However, programs like the Positive Behavior Intervention Supports-*Plus* (PBIS*Plus*) coaching model and the Classroom Check-Up (CCU, Reinke, 2006) classroom coaching have both demonstrated success in supporting teacher's implementation of a variety of evidence-based classroom management practices (Pas et al., 2013). Still, more research is needed to investigate the efficacy of coaching in improving teachers' general CM practices outside of these programs.

Like coaching, there is a substantial body of literature that supports the use of PF for increasing teachers' use of praise and implementation of interventions. Compared to traditional methods of PD like workshops and yearly trainings, performance feedback is much more effective in facilitating long-term improvement in teachers instructional and behavioral practices in the classroom (Elmore, 2002; Sugai et al., 2000). Promoting teacher skill development requires systematic and continuous teacher support practices like performance feedback (Sugai et al., 2000). The individualized and on-going nature of performance feedback allows for a more targeted approach to PD that provides teachers follow-up support after the initial training or workshop. Performance feedback was also shown to be effective even without any substantial initial training for teachers which reduces the time investment for teachers (Noell et al., 2005).

Cavanaugh (2013) conducted a literature review of studies testing the effect of performance feedback on teachers' use of praise and opportunities to respond (OTR). Generally, performance feedback was shown to be effective at increasing teachers use of praise and, less consistently, OTRs. Studies included in the review tested a range of grade

levels from preschool to high school and in both general education and special education classrooms (Cavanaugh, 2013). Without some type of structured follow-up like performance feedback teacher's implementation of intervention plans was found to be significantly lower than with it (Noell et al., 2005). There are a variety of different methods for delivering performance feedback including self-monitoring of audio or video, presenting graphs, and e-mails describing teachers use of effective practice (Cavanaugh, 2013). When used together these different methods were more effective than the use of a single method alone (Cavanaugh, 2013). Additionally, for teachers who did not benefit from performance feedback alone, setting goals was an effective practice when used in conjunction with performance feedback (Kalis et al., 2007). There is a plethora of support for performance feedback in the literature, but it primarily focuses on increasing teachers use of praise and/or OTRs. More research is needed to examine the utility of performance feedback in improving teachers' general CM practices.

Coaching and PF Implementation Considerations

A strong evidence base (i.e., evidence of efficacy or effectiveness) should be a prominent and primary consideration for utilization or adoption of any school-based policy, procedure, practice, service, support, or intervention. Unfortunately, effectiveness is only one of numerous factors that influence school-based practice use (Timperley et al., 2007). Some of the most frequently cited considerations are time and resource cost, accessibility, acceptability, feasibility, use of outside expertise, theoretical and research base, contents, mode of delivery, role of local leadership, and systematic follow up and

feedback (Timperley et al. 2007). Such factors are considered when schools adopt PD activities including coaching and PF.

Cost. Traditional on-site coaching programs are a resource-intensive intervention (Kraft et al., 2018). Knight (2012) did a cost analysis of coaching across several schools and found that costs ranged from $3,300 to $5,200 per teacher. Another prominent consideration for coaching as a PD activity is that the positive effects may not be consistently generalizable (i.e., used on a larger scale; Kraft et al., 2018). For example, Kraft et al. (2018) found that larger programs only saw a fraction of the positive effects of coaching that were observed in smaller programs, meaning that coaching is best implemented as a smaller scale targeted program. There are several possible reasons for this including the challenge of building a large enough staff of well-trained effective coaches and achieving teacher buy-in across larger populations (Kraft et al., 2018). The current literature suggests that coaching has not been successfully implemented on a larger scale in a high-quality and cost-effective way (Kraft et al., 2018). Additionally, performance feedback is usually a time and resource intensive practice which requires frequent observations and data collection by a staff member, analysis and presentation of data, and regular meetings with the teacher to provide feedback.

Accessibility. Coaching and PF PD activities are not always accessible to educators as they require an external observer to collect data and/or provide feedback. Typically, PD takes the form of in-service trainings/workshops that are provided to all teachers regardless of their specific needs/areas of improvement (Bayar, 2014; Corcoran, 1995; Parise & Spillane, 2010). Because teachers are not able to access and implement

these PD activities without external support, they are less accessible than strategies that teachers can use independently.

Acceptability. Though not widely researched, there are several studies that examined the acceptability of coaching and/or performance feedback as a professional development activity (Pas et al., 2016; Strohmeier et al., 2014). Strohmeier and colleagues (2014) examined the social validity of different training methods used to improve performance. This study found that performance feedback was the most acceptable and highly rated training intervention for improving implementation of treatments (Strohmeier et al., 2014). Additionally, acceptability was linked to the effectiveness of the methods used for improving treatment integrity Strohmeier et al., 2014). In regard to coaching, Pas and colleagues found that over 90 percent of teachers indicated a high level of acceptability for the Classroom Check-Up (CCU) coaching model (Pas et al., 2016). However, 25 percent of teachers indicated concerns about the amount of time required to participate in the coaching (Pas et al., 2016).

Feasibility. In most studies on performance feedback, researchers or university students who often have specific training in data collection procedures, intervention implementation, and consultation, are the ones providing feedback to teachers (Cavanaugh, 2013). Additionally, researchers and university students often have more availability to work with teachers and provide performance feedback than regular school staff might have.

Self-monitoring

While generally accepted as effective, coaching and PF may not be a feasible PD choice for some buildings or districts given their potential time, resources, and monetary costs associated with these approaches. One potentially more feasible PD option that incorporates elements of coaching and PF is self-monitoring. Self-monitoring is a strategy for managing or regulating one's own behavior (Bruhn et al., 2015). Self-monitoring often includes observing, evaluating, and recording a person's own behavior(s). The behavior(s) being targeted should be explicitly defined prior to the start of the self-monitoring activity. The goal of self-monitoring is for the individual to achieve independence with a strategy or skill and maintain positive outcomes without external support (Bruhn et al., 2015).

Simonsen (2020) emphasized the utility of self-management (which includes self-monitoring) as a realistic, efficient, and effective approach to supporting teachers' implementation of empirically supported practices in the classroom. Research has demonstrated the effectiveness of self-monitoring at changing behaviors in a variety of settings and with a variety of participants (Kalis et al., 2010). However, there is limited research on teacher self-monitoring behavior (Kalis et al., 2010).

SM Effectiveness

Broadly, SM is recognized as an effective practice. Rispoli and colleagues (2017) conducted a systematic review of teacher self-monitoring interventions and found that self-monitoring generally led to positive changes in teacher behavior. However, most of the studies included in this review were limited to increasing teachers' use of praise in the

classroom (Rispoli et al., 2017). The majority of studies required teachers to track their behavior during each observation period using event recording techniques (e.g., frequency counts). Interestingly, the accuracy of teachers' self-monitoring data did not impact the effectiveness of the intervention, which suggests the simple act of monitoring one's own behavior can lead to a change in behavior (Rispoli et al., 2017). The results of this review demonstrate that teacher self-monitoring has been used successfully to change specific behavior practices in the classroom. Further research is needed to evaluate the effectiveness of teacher self-monitoring on improving general CM practices and without the use of event recording.

Although SM research in educational settings is sparce, a greater body of literature supports SM generally, with a large number of studies examining self-monitoring interventions for students that have found positive results. For example, Guzman and colleagues (2018) conducted a meta-analysis of self-monitoring interventions on reading performance in K-12 students. Results across 67 participants included in the meta-analysis indicated that self-monitoring had a significant positive effect on the reading performance of K-12 students and is therefore an evidence-based reading intervention (Guzman et al., 2018). Another article by Hallahan et al. (1993) reviewed studies on self-monitoring and concluded that self-monitoring of attention during academic work improves on-task behavior and can lead to increases in academic productivity. Other studies have found positive effects of different self-monitoring interventions on students with learning disabilities, attention-deficit hyperactivity disorder (ADHD), and autism spectrum disorder (ASD) (Harris et al., 2005; Holifield et

al., 2010; Reid & Harris, 1993). These studies and many others have demonstrated that self-monitoring of attention, performance, and procedure can all be effective at improving different aspects of classroom behavior and academic performance (Harris et al., 2005; Holifield et al., 2010; Lan, 1996; Reid 1996; Reid & Harris, 1993; Rock & Thead, 2007).

SM Implementation Considerations

There are decades of research supporting the use of self-monitoring in the classroom as an intervention for both students and teachers. Self-monitoring has proven to be effective for students in various settings (e.g., general education and special education classrooms) and in various areas of academics (e.g., math, reading, spelling; Menzies et al., 2008). Self-monitoring is an easy to implement, flexible, and effective intervention that can be customized to the individuals' specific needs. For example, self-monitoring of attention, academic performance, and behavior have all been used effectively to improve specific aspects of students' academic performance and/or classroom behavior. Additionally, self-monitoring requires minimal time and resources to implement effectively because the individual that is being targeted for the intervention is also the person collecting data and observing their own behavior. The individual engaging in self-monitoring typically requires minimal training to be able to benefit from the intervention.

Although self-monitoring has many advantages it is not without its own set of challenges. First, although self-monitoring is a customizable intervention, it is often not personalized to an individuals' strengths and weaknesses. In order for self-monitoring to

have the most potential for improving behavior and academic performance, tailoring monitoring to the individual is recommended (Dunlap et al., 1995; Menzies et al., 2008). Next, training is a critical part of the self-monitoring intervention because if the individual has not mastered the procedures of the intervention and does not implement it with fidelity, it has a lower chance of success (Vanderbilt, 2005). Self-monitoring also places more of the burden on the individual which can be especially problematic for teachers who may not be able to consistently engage in the intervention during teaching. Finally, self-monitoring may not have sustained effects or generalize well to other settings and skills as there is limited research to support long term benefits of the intervention (Menzies et al., 2008)

Multi-Tiered System of Educator Support

Due to the importance of PD for improving CM and the challenges of implementing it effectively, a framework designed to meet the various support needs of all teachers is necessary (Grasley-Boy et al., 2021). Given the wide range of teacher knowledge, skills, and abilities, classroom educators will need varying levels of support regarding improving CM practices and engaging in PD activities (Grasley-Boy et al., 2019). Additionally, teacher professional development needs are not limited to CM. As such, consistent with support efforts for students, a potential solution by which to flexibly support teacher PD is seen in the use of a multi-tiered support framework. Within a multi-tiered framework, a continuum of PD activities is organized by time and resource intensity to correspond with level of support needed (i.e., Tier 1, 2, and 3). This continuum of PD activities builds on current PD options like workshops but adds more

intensive options for teachers who require more support. Additionally, PD provision decisions within a multitiered PD framework (i.e., matching PD supports to teacher need) should be data-based, or driven by frequent observations, screening, and progress monitoring assessment activities. Ideally, data collection begins with screening and regular progress monitoring to assess response to the current level of support being provided (Grasley-Boy et al., 2019). While MTSS frameworks are prominent in student support efforts, the use of MTSS to support educator PD is relatively novel. Recently however, scholars have noted the promise of such an approach and have increasingly called for its application to support teacher PD activities, including those targeting CM practices (see Sims et al., 2021; Gage et al., 2017; Gage et al., 2018; Macsuga-Gage, 2013; Myers et al., 2011; Simonsen et al., 2016; Grasley-Boy et al., 2019). To date, research evaluating multitiered PD efforts remains sparce, but appears promising for improving teachers' CM practices (Gage et al., 2017; Gage et al., 2018; Macsuga-Gage, 2013; Myers et al., 2011; Simonsen et al., 2016). Despite noted shortcomings in pre-service training and traditional, workshop type PD approaches, multitiered PD frameworks have yet to be widely used is educational settings. Unfortunately, traditional PD activities (i.e., workshops, in-service trainings, etc.) still dominate the field with little lasting positive effect on teacher's CM practices.

Tiered CM PD Activities

While current, typical in-service PD activities (i.e., workshops, lecture-style trainings) have not demonstrated noteworthy, durable positive impacts on teacher CM practices, more intensive, ongoing, data-driven activities (e.g., programs, Coaching, PF)

have demonstrated more success in improving CM practices (Duchaine et al., 2011; Myers et al., 2011; Pisacreta et al., 2011). Within a Multitiered System of Educator Support (MTSES; Sims et al., 2021) the former, typical in-service workshops, would mostly likely be considered a Tier I support as it would likely be provided for everyone. Currently, these more productive PD activities (e.g., performance feedback, coaching, consultation) are often embedded within a comprehensive, multicomponent PD approach (e.g., CHAMPS cite; Incredible Years cite) and address many of the limitations of in-service/workshop type PD activities. While effective, a comprehensive, multicomponent, and curricular PD approach which includes coaching and performance feedback elements, is often both time and resource intensive, which would situate these activities at the Tier III level. To this point, Simonsen (2020) found that coaching is 6 to 12 times more costly than typical workshop-type in-service PD activities. Schools often have limited resources to dedicate to PD activities which means that a comprehensive, multicomponent, curricular PD may not be feasible for many districts/buildings. Given multicomponent, curricular PD programs utilizing coaching and PF (i.e., Tier III) may be too resource intensive (e.g., cost, time, training) and in-service trainings (i.e., Tier I) appear less than effective, development and evaluation of additional, less intensive PD activities (i.e., Tier II) is necessary.

Within an MTSES framework, Tier I supports would be provided to all teachers and would likely continue to take the form of one time, presentation style workshops and trainings. The goal of these activities could be informational or preventative. Topics covered in the workshops and trainings could or could not be matched to teachers' areas

of need or interest. A large amount of information is condensed into a single training, which makes it difficult for many teachers to remember and apply the information gained when they get back in the classroom (Hunzicker, 2011). This type of PD is the most widely used form of PD in schools though it is not effective for all teachers. While some teachers can benefit from these large-group trainings on a specific skill or skills, others need more targeted follow-up support to successfully implement the skills learned.

For teachers that need more support than Tier I large group workshops, other more intensive PD options exist. There is now a growing body of literature supporting the use of coaching and performance feedback to facilitate teachers' development of specific skills. Typically, the coaching or performance feedback intervention will target one specific CM skill for improvement based on the teachers' specific needs. These PD opportunities would be considered Tier III interventions because they are individualized interventions that are very time and resource intensive. Coaching and performance feedback involve frequent observations, data-collection, meetings, and discussions with the coach/observer and the teacher. While generally effective, coaching and performance feedback are not feasible or sustainable practices on a large scale or for more than a small number of teachers.

Expanding CM PD Options

Some teachers PD needs will be met by Tier I efforts, others will likely need some PD support beyond these large group workshops. Consultee-centered consultation is considered a Tier II intervention because it is less time and resource intensive than coaching and performance feedback, but it involves direct support from another

professional. Consultation often uses a problem-solving approach to work through challenges with a particular student that the teacher has sought consultation for. Focusing on problems with a particular student can lead to attributing difficulties to within a student rather than focusing on the teacher and the instructional environment (Reinke et al., 2008). Unfortunately, there are limited options for Tier II supports beyond consultation which does not emphasize teacher skill development and improving general CM skills.

Importance of Data in MTSES

As noted previously, assessment is a critical part of MTSES, as it is the foundation of data-based decision making (Sims et al., 2021). It provides researchers with data regarding current performance and progress and allows for informed, assessment driven, objective decisions about the supports and interventions that are necessary (Linan-Thompson et al., 2022). One important aspect of data collection within a multitiered PD framework is using a screening tool to determine which teachers could benefit from additional support beyond Tier I trainings. When collecting assessment data for use within a multi-tiered PD system to provide performance feedback, the assessment data is essentially the intervention making it especially important to select a technically sound instrument (i.e., valid, reliable, diagnostically accurate). Using technically sound instruments that are valid and reliable for the target population is vital in ensuring that the decisions made are based on accurate data.

Additionally, one of the most important components of MTSES, and coaching, performance feedback, and consultative approaches to PD in particular, is data-based

discussions about current performance and progress (Reinke et al. 2008). In coaching and performance feedback, data regarding teachers' current performance is collected daily or multiple times per week during regular observations by a coach/observer. This data is then compiled and analyzed for discussion and decision-making regarding the intervention. Typically, teachers would receive coaching/performance feedback several times a week which can include specific verbal and visual feedback, prompting, and a discussion of performance based on the data collected. Psychometrically sound data helps to determine where support is needed, what type of support is needed, and evaluation of the supports themselves (Sims et al., 2021). Presenting and discussing data with teachers is a critical part of the intervention for coaching and performance feedback as it informs every part of the process from start to finish. For the uses described above, choosing a data collection method that provides accurate and reliable data without being too time and resource intensive is critical. Unfortunately, the availability of such instruments for measuring CM behaviors is very limited (Sims, 2016; Reddy et al., 2013).

CM Assessment

Despite its importance, historically, the availability of assessments for assessing educator use of evidence-based classroom management (CM) practices has been limited. Currently available assessments (e.g., MOOSES; CLASS) may be too costly, complex, time-intensive, or difficult to access. There is a significant difference between how teacher use of CM practices is assessed in schools and research. Typically, school administrators (e.g., principals) are responsible for evaluating teachers, with specific guidelines and procedures varying by district (Rigby, 2014). These evaluations are

typically limited to one 30 to 60-minute classroom observation per teacher, where the administrator observes the teacher's performance and fills in an observation rubric or takes field notes (Reid, 2020). However, ambiguity in the evaluation process and subjective ratings prone to bias are common issues, with some districts providing little guidance or definition of evaluation criteria (Reid, 2020; Rugby, 2015). Consequently, the lack of consistency and objectivity in these evaluations diminishes their value.

When evaluating teachers' CM practices in research, systematic direct observation (SDO) is often used to collect performance data and is considered the gold standard of behavior assessment measures because it provides reliable, accurate measurements of target behavior (Riley-Tillman et al., 2011). In SDO, an observer records some specific aspect (frequency, duration, latency, etc.) of an operationally defined behavior during, not after, the observation. Though this method of measuring behavior yields accurate and reliable results, it is not the most feasible way to collect data. SDO can be time and resource intensive due to the extensive training, multiple observations, and consistent focused attention required to collect data (Riley-Tillman et al., 2011). With SDO, typically only one specific aspect of behavior is measured which makes it difficult to gather data on a set of behaviors (i.e., CM behavior) at once.

There are advantages of SDO which make it an excellent option, but it is not suited for every situation. It is important to consider a variety of factors when selecting a data collection procedure including the purpose and intended use of the assessment data as well as the time and resources required. The assessment tools most frequently seen in CM research are often costly, complex, time intensive, training intensive, or are unable to

capture the more nuanced aspects of CM. The Direct Behavior Rating-Classroom Management (DBR-CM; Sims et al., 2021) was developed to address the limited availability of feasible, reliable, and accurate CM assessment.

DBR-CM Background

Development

Sims et al. (2021) created DBR-CM as an assessment tool to complement the existing CM assessment options. The DBR methodology combines the strengths of systematic direct observation and behavior rating scales, while overcoming some of the barriers to their use (Chafouleas et al., 2021). This results in a practical, adaptable, and defensible assessment tool that can be used to evaluate various classroom behaviors, including teacher CM behaviors. According to Kane's (2013a) argument-based validation model, development and validation of an assessment begins with a statement of the intended interpretations and uses of assessment scores. The interpretation and use argument (IUA) for the DBR-CM stated the observation-based assessment would efficiently generate feasible, defensible, flexible, and usable data for screening and formative assessment of educator CM practices (Sims et al., 2021). The development process involved a comprehensive literature review to identify evidence-based CM practices, content validation activities to create items based on similar CM practices, developing operational definitions for each item, and organizing items into a scoring format used by DBR Single Item Scales (DBR SIS; Sims et al., 2021).

Direct Behavior Rating-Single Item Scales. Research shows that DBRs, which have primarily focused on student behaviors, are a psychometrically sound, adaptable, and

practical method for assessing behavior (Chafouleas, 2011). DBRs merge the strengths of general outcome measures and rating scales into a low inference assessment format that reduces the time between observations and ratings, making them suitable for efficient screening and formative assessment data collection (Chafouleas, 2011). Advantages of using DBRs include their ability to generate objective and reliable scores with limited attention and training required, their utility for interventions such as PF and self-monitoring, and their reduced strain on resources like time and cost (Christ et al., 2009; Harrison et al., 2014). In summary, the DBR assessment methodology has several strengths that are well-suited for its intended uses in DBR-CM (i.e., screening, formative assessment, and direct support).

Validation

Arguments-based Approach to Development and Validation. Validation of the DBR-CM assessment tool involves a series of studies designed to test the intended interpretations and uses of the generated scores. The validation process includes five interrelated inferences that connect instrumentation to scores to score interpretation and use: scoring inferences, generalization inferences, extrapolation inferences, decisional inferences, and theory-based inferences (Kane, 2013a, 2013b). The validation process involves collecting and analyzing psychometric reliability and validity evidence using representative samples of all intended assessment subjects. The assessment is considered valid when the accumulated psychometric evidence supporting the IUA claims outweighs any counterclaims (Kane, 2013b).

Preliminary DBR-CM Validation Work. In an initial study conducted on a sample of 107 elementary school classrooms, significant positive correlations were found between DBR-CM scores and concurrent measures of behaviors and characteristics of classroom instructional environments, such as the Classroom Atmosphere Scale, Brief Classroom Interaction Observation-Revised, and Ohio Teacher Self-efficacy Scale. The study also found significant correlations, in the expected direction, between DBR-CM item ratings and SDO variables that measure teacher CM behaviors, such as Praise, Opportunities to Respond, and Reprimands.

Furthermore, the study showed that measures of inter-rater reliability (IRR) exceeded acceptable levels for all but one item (Praise), even in the absence of formal DBR-CM training. While these findings are promising, the psychometric evidence provided by this initial study is not sufficient to support the use of DBR-CM across all proposed interpretations and uses. Further validation work is needed to strengthen the evidence base for the DBR-CM.

DBR-CM Supported PD

As noted previously, current approaches to PD center around large group lecture style workshops and trainings which are not effective for all teachers. Beyond Tier I, workshop trainings, much of the current literature focuses on comprehensive, Tier III PD programs incorporating, but not limited to coaching and performance feedback, which may be too time and resource intensive to be feasible in schools. Performance feedback intervention research generally examines more time and resource intensive methods of delivering performance feedback (Reinke et al., 2008; Reinke et al., 2012; Sprick, 2013).

Though these interventions have documented positive impact on teacher CM behavior, this type of performance feedback intervention has limited practical use in the schools due to the lack of time and resources available for PD activities. The DBR-CM could prove useful in a less intensive performance feedback intervention, but there is limited research regarding its utility in this context. Previous research that utilized the DBR-CM for a performance feedback intervention had limited success which calls for further research on this topic.

Prior DBR-CM PD Work

Previous work in this area yielded mixed results. Sims (2016) evaluated the effects of a quick, efficient, and simple procedure that provided teachers with regular PF using the DBR-CM. Teachers were observed for a brief period (i.e., 15 minutes), daily for several weeks. Following each observation, participating teachers were presented with a graph of their DBR-CM scores. To limit PD activity intensity (i.e., utilization of time and resources), this simple procedure allowed participating teachers to ask questions, but did not involve lengthy meetings, coaching, prompting, or qualitative feedback regarding ratings across each area of performance (i.e., Praise, Communication, Engagement, Enthusiasm, Rapport). During the performance feedback intervention, teachers were provided with a graphic representation of their DBR-CM scores following an observation by an external observer. A multiple baseline design across five participants over 45 days was utilized for this study. Results indicated that two of five teachers showed improvement in CM behaviors during the performance feedback intervention. Sims

(2016) discusses several factors that could help explain the limited success of the intervention and directions for future research that will be addressed in the present study.

Sims (2016) noted that the timing of the intervention and observations is important for several reasons. First, planning the intervention for a time of year that would allow for data to be consistently collected without interruptions from major events in the school year (e.g., holidays, spring break, etc.) will limit the influence of these events on study results. Intentional planning of external observations is also important for obtaining accurate DBR-CM scores. Observations should occur during teacher guided instruction or teacher-student interaction when teachers have the greatest opportunity to engage in CM behaviors allowing for accurate measures of these behaviors.

Sims (2016) also suggested modifications to the procedures of the intervention that might lead to more desirable results. First, rather than presenting performance feedback after each observation period, it may be beneficial to do so before the start of each observation to tap into possible priming effects. Second, the addition of a formal training component for teachers to increase their understanding of the CM behaviors that are being measured could support the use of positive CM behavior. Third, expanding performance feedback from visual presentation of results only to including further elaboration or interpretation of observation results could allow teachers to better understand the feedback they receive (Sims, 2016). However, careful consideration must be made to maintain the minimally time and resource intensive aspect of the intervention to protect the purpose of the study. The present study will replicate prior application of the DBR-CM to drive CM PD efforts. This replication will utilize this simple

performance feedback procedure first attempted by Sims (2016) with minor changes to address limitations noted in prior work as well as the addition of teachers own assessment of their performance via a self-monitoring component. The present study attempts to addresses these limitations and seeks to improving future applications of the DBR-CM within performance feedback interventions. Goals of this work are threefold, as it will address several gaps in the literature including limited research on (a) performance feedback interventions using the DBR-CM, (b) less intensive performance feedback interventions, and (c) the effectiveness of combining performance feedback with self-monitoring.

Study Purpose

The purpose of this study is to evaluate the effect of a performance feedback and self-monitoring intervention on teacher's CM behavior. Teachers' classroom management practices will be observed and rated using the DBR-CM ER and then they will receive both visually and verbally presented feedback prior to the start of the next observation session. Teachers will also engage in a combined performance feedback with self-monitoring PD activity using the DBR-CM SR during the second portion of the study. This study is guided by four primary research hypotheses.

1. Performance feedback-based PD will increase teacher use of evidence-based CM practices.
2. Combined performance feedback with self-monitoring PD will increase teacher use of evidence-based CM practices.

3. Teachers will rate DBR-CM performance feedback as an acceptable PD practice.

4. Teachers will rate the use of DBR-CM performance feedback and self-monitoring as an acceptable PD practice.

Methods

Participants

Participants included four general education teachers from a public elementary school in Florida. To be included in the study, participants had to currently be teaching in a general education setting and be credentialed by the state Department of Education. Additionally, participating teachers had to display mild to moderate deficits in their use of CM practices to be included in the study. A mild to moderate deficit was defined as a median DBR-CM ER Total score between 12 and 28 (i.e., average item score between 3 and 7). A total of five potential participants were observed three times by an external rater to ascertain their median DBR-CM ER Total score as part of the screening and inclusion process. One teacher obtained a median total score of 33 during screening and was excluded from the study. Therefore, a total of four participants met inclusion criteria and were retained for this study.

Measures

Demographic Survey

A study specific survey was used to collect general demographic information for participants. Demographic information collected included gender, race/ethnicity, highest degree completed, area of certification, grade taught, years of experience, and number of

students in a class. The demographic survey included forced choice and open response questions and was completed via the web-based survey program Qualtrics.

Direct Behavior Rating –CM

External Rater Form. The Direct Behavior Rating-CM (DBR-CM) assesses teacher use of evidence-based CM practices (Sims et al., 2021). The DBR-CM external rater (ER) form is completed by an outside observer who rates a teacher's CM behavior directly following a specified observation period during instructional activities. To facilitate efficient completion of ratings and in attempt to be more environmentally conscious and resource frugal, DBR-CM ER and SR ratings were completed via a Qualtrics survey formatted DBR-CM form. Teacher use of CM practices is measured in four domains of CM praise, communication, enthusiasm, and rapport. Users are tasked with rating the amount or level of CM behavior associated with each item using a Likert scale that ranges from 0 = *Low* to 10 = *High*. Higher scores indicate greater amounts or levels of CM practice use (Sims et al., 2021). Individual item ratings can be added together to calculate a DBR-CM Total Score, or score indicative of a teacher's overall use of CM practices. The DBR-CM has been shown to have acceptable levels of inter-rater reliability as measured by Intraclass Correlation Coefficient (ICC). ICC values ranged from .67 to .84 (Sims et al., 2021). Additionally, DBR-CM scores were found to have significant correlations with overall scores (i.e., BCIO-R, CARS, OSTES; r = .25 to .81) and individual scores (r = .41 to .78) on concurrently completed measures (Sims et al., 2020).

Self-Report Form. Like the DBR-CM ER, the self-report (SR) form is designed to measure teacher use of evidence-based CM practices. The DBR-CM SR is completed by the teacher to assess their own use of evidence-based CM practices (Sims, 2022) during a specified observation period during instructional activities. The DBR-CM SR is largely consistent with the DBR-CM ER, with minor grammatical and semantic differences in instructions, item definitions, and examples. Participants completed DBR-CM SR forms via a Qualtrics link that was sent via text and/or email after each observation.

Modified User Rating Profile-Intervention Revised

The User Rating Profile-Intervention Revised (URP-IR; Chafouleas et al., 2011) assesses user perceptions of acceptability, understanding, family-school collaboration, feasibility, system climate, and system support of an intervention (Briesch et al., 2013). This measure is meant to provide information about facilitators and barriers to usage of an intervention that exist at different levels (i.e., individual, intervention, and environment). Teachers were asked to rate performance feedback and self-monitoring and performance feedback as two separate interventions. The URP-IR uses 29-items to assess rater perceptions across these six dimensions. Items are completed using a Likert scale ranging from strongly disagree (1) to strongly agree (6). Higher scores, after reverse scoring for some items, are indicative of more favorable user perceptions within a given dimension. Rater responses are aggregated to provide a metric of rater perception across respective domains as well as an overall usage rating score. Given the focus of this study (teacher professional development), participants completed a modified version of the

URP-IR via Qualtrics which excluded items comprising the family-school collaboration subscale.

Fidelity Measure

Fidelity was monitored across adherence (i.e., number of intervention components completed), quality (e.g., enthusiasm implemented), and duration (i.e., number of treatments completed; time to complete each treatment). A study specific measure of implementation fidelity was completed by the observer daily during treatment phases (i.e., Phases B and C). Quality was measured using a brief Likert-type questionnaire which was completed by teachers after each performance feedback session in both Phase B and C. Adherence measures varied slightly across phases, with Phase C (i.e., performance feedback with self-monitoring) including items addressing self-monitoring components (e.g., Teacher completed DBR-CM SR ratings).

Design

This study utilized a combined multiple baseline, ABC single case design across participants to evaluate the impact of performance feedback and performance feedback plus self-monitoring PD activities with four volunteer participants. SCD are particularly well-suited for applied settings, as interventions typically focus on individual subjects and many of the requirements of between-groups experimental designs (e.g., randomly assigning subjects to a treatment group, standardizing treatments among subjects, etc.) are not feasible (Kazdin, 1982). Thus, single-case designs provide a solution that allows researchers to conduct experimental investigations to evaluate a treatment or intervention with only a small number of subjects, as is common in a school setting (Kazdin, 1982).

Multiple baseline designs are characterized by staggering the introduction of a treatment (i.e., independent variable) chronologically across participants (Kratochwill et al., 2010). This is to say observations of baseline performance of the same behavior for each person are conducted until the behavior stabilizes (Kazdin, 1982). After a stable rate of baseline behavior is reached for each participant, one participant moves into the intervention phase while others continue in baseline. This process is repeated until eventually all participants are receiving the intervention. The behavior of the person receiving the intervention is expected to start changing from baseline at the time point when the intervention is implemented. Because each person is exposed to the intervention at a different point in time, there are multiple opportunities to observe this change during the intervention phase, while controlling for maturation effects.

The strengths of this design lie in its ability to allow all participants to receive intervention, while maintaining ability to establish 3 affirmations of the consequence (i.e., effect and subsequent replication of effect). Limitations of this design may include difficulty establishing a stable baseline, delays in beginning treatment for some participants, and practical limitations like duration of the study (Horner et al., 2005; Kazdin, 2011; Tawney & Gast, 1984).

What Works Clearinghouse (WWC) assembled a panel of experts in single-case design (SCD) to create a set of SCD standards. WWC provides standards for all SCD research as well as more specific standards for each of the different types of single case design. There are three ratings for WWC SCD studies: *Meets WWC SCD Standards*

Without Reservations, Meets WWC SCD Standards With Reservations, or Does Not Meet WWC SCD Standards.

For a SCD study to *Meet WWC SCD Standards Without Reservations,* specific design criteria must be present. These criteria are organized under the following categories: independent variable, interassessor agreement, attempts to demonstrate effect over time and data points per phase, and other concerns. Regarding the independent variable, raw data must be made available to allow for visual analysis which helps the WWC assess whether WWC standards are met. WWC standards also require that the independent variable be systematically manipulated meaning that the researcher will choose how and when conditions change. The outcome variable must be measured systematically over time by multiple assessors so that interassessor agreement (IAA) can be calculated. IAA data must be collected in each phase for a minimum of 20 percent of the data points in each condition (i.e., baseline and interventions). Minimum acceptable values of IAA depend on the method used to calculate it. Minimum values are 0.80 for percentage agreement and 0.60 for Cohen's kappa. These values must be met for each outcome across all phases and cases to meet WWC standards. Other concerns include confounding factors (which must not be present) and no overlap of training phases. It is important for studies to avoid overlap in the training phases among subjects in the experiment so that threats to internal validity may be excluded. Finally, the study must include a minimum of three attempts to demonstrate an intervention effect at three different time points. Criteria for the number of data points and phases required to meet standards are specific to design type. The present study utilized a multiple baseline design

which must have a minimum of six phases with five data points per phase to meet WWC standards without reservations. Additionally, the implementation of the design must involve a degree of concurrence to set it apart from reversal/withdrawal (AB) designs.

ABC designs across participants involve two independent variables (B and C). First, observations of baseline performance (A) of the same behavior for each person are conducted until the behavior stabilizes, then the first participant begins the intervention (B), next the second participant begins the intervention while the first participant moves to the second intervention (C) (Kazdin, 1982). After a stable rate of behavior is reached for each participant in each phase, one participant moves into the intervention phase while others either continue in baseline or enter a different intervention phase. This process is repeated until eventually all participants are receiving the second intervention.

Dependent Variable

DBR-CM Total Score. Teacher CM was measured using the DBR-CM ER, which is made up of four constructs that combine to rate teacher's overall CM including praise, communication, enthusiasm, and rapport. Praise is 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" (Sims, 2016, p. 39). Communication is "clearly conveying goals and expectations of a classroom and/or instruction period to students" (Sims, 2016, p.39). Enthusiasm is defined as "the delivery of instructional content in a meaningful, memorable, and/or engaging manner" (Sims, 2016, p. 41). Rapport is defined as "the quality of the student-teacher relationship especially that of mutual trust, emotional affinity, acceptance, and positivity" (Sims,

2016, p. 40). Examples and more detailed definitions of each construct are provided on the DBR-CM ER form (see appendix C). Individual item scores range from 0 (Low) to 10 (High). DBR-CM ER Total Scores range from 0 to 40. Higher scores are more favorable as they reflect that more positive CM behaviors are present. DBR-CM ER scores in each of the four constructs were displayed in bar graph format for every observation in each intervention phase. Scores from each observation session were added to one graph so that progression and performance over time could be observed.

Procedure

Data Collection

Training. Prior to the start of data collection, external observers completed an approximately 60-minute training and reliability check to facilitate reliable completion of DBR-CM ER ratings. Initially, data collectors completed a 40-minute, web-based training. This training presents the background and rationale for the development of the DBR-CM as well as the importance of the use of effective, evidence-based CM practices. Additionally, the training introduces and defines DBR-CM items and provides examples and nonexamples associated with each. This training also covers item completion, including an introduction to the response format and Likert scale, as well as more detailed descriptions of the scale anchors that guide ratings. Trainees are given several opportunities to practice by watching and rating video examples of teacher CM. Lastly, data collectors completed a reliability check prior to collecting any study data. The reliability checks required data collectors to rate three different teaching video clips using

the DBR-CM ER form. Study personnel were required to reach a percent agreement level of 90 percent or better across the three video examples.

Before entering the second treatment phase (i.e., Phase C), teacher participants also completed the web-based training and reliability check processes for the DBR-CM. Teacher participants completed this training to facilitate completion of the DBR-CM SR during the performance feedback plus self-monitoring PD activities. While not an outcome measure for the study, reasonably informed and reliable self-evaluation (i.e., completion of DBR-CM SR ratings) is an important mechanism within the PD activities.

Participant Screening. To facilitate study participant inclusion, direct observation took place three times for 20 minutes using the DBR-CM ER. Observations occurred on separate days for each of the respective participants. From these observations, the mean DBR-CM ER Total Score was used to determine if a potential participant displayed mild to moderate deficits in their use of CM practices (i.e., a DBR-CM ER Total score between 12 and 28). The rationale for this selection criteria is multifaceted. First, broadly, DBR assessment methodology has established ratings between 3 and 7 to be indicative of mild to moderate performance deficits. This may also be thought of as ratings from 0 to 2 typically indicate perceptions of significant performance deficits, whereas ratings from 8-10 indicate perceptions of high-performance levels (Chafouleas et al., 2015) Mild to moderate deficits in use of evidence-based CM practices allows room for improvement in participant behavior. While more pronounced performance deficits (i.e., average item ratings 0 to 2) allow for participant improvement, in this instance, such ratings likely represent a mismatch between issue

(i.e., deficient performance) and solution (i.e., PD activity). To frame this consideration in a potentially more familiar multi-tiered service delivery framework perspective, the PD activities utilized in this study are consistent with low intensity, minimally invasive treatments (i.e., Tier II). Participants with DBR-CM Total Scores below 12 would be consistent with items scores between 0 and 2 and would be indicative of severe performance deficits that would require more intensive, invasive treatments (i.e., Tier III). Five participants were screened for inclusion, but only four participants yielded an average total score between 12 and 28 and were retained for study participation. The fifth teacher that was excluded from participating in the study earned an average DBR-CM total score of 33 across the three initial observations.

In-Study Data Collection. Ongoing data collection occurred using external observer completed DBR-CM ER forms to evaluate the effects of the two study treatments. The principal investigator worked with participating teachers to identify core instructional blocks, or periods of 20 minutes or more in which the teacher was engaged in math, reading, writing, science, or history instruction. Further, the principal investigator coordinated observation times for instructional periods that corresponded to external observer availability whenever possible. The principal investigator also ensured that external raters and participants had access to appropriate links to DBR-CM ER and SR Qualtrics survey links.

Baseline. To evaluate pretreatment (i.e., PD activity) amount or levels of teacher CM practice use, a minimum of five 20-minute baseline data (i.e., Phase A) observations occurred on separate days. External observers took mental note of teachers' use of CM

practices throughout their observations. Following each 20-minute observation period, data collectors completed ratings of participant's amounts or levels of CM practices using the DBR-CM ER form. Ratings occur across the four individual DBR-CM ER items (i.e., Praise, Communication, Enthusiasm, Rapport), which facilitates calculation of a DBR-CM Total Score (i.e., the primary study outcome variable). To evaluate the reliability of baseline data, per recommendations by What Works Clearinghouse (2022), 20 percent of baseline phase observations were scheduled to include a second external observer. The inclusion of a second rater allowed for the evaluation of interobserver agreement levels.

Treatment Phases. To evaluate the amounts or levels of teacher CM practice use during PD activities (i.e., Phases B, C), external observers completed 20-minute observations daily (i.e., once each day). Consistent with baseline data collection procedures, external observers took note of the amounts or level of teacher CM practice use throughout their observations. At the conclusion of each 20-minute observation period, data collectors completed ratings of the amounts or levels of teacher use of CM practices using the DBR-CM ER. Observations were completed by the primary investigator. A second rater was scheduled to conduct a concurrent observation for 20 percent of treatment phase data to evaluate levels of interobserver agreement.

Interobserver agreement. To ensure reliability of study data, 20 percent of observations across all phases were scheduled to include two external observers, who would complete concurrent observations and DBR-CM ER ratings. IOA levels were evaluated weekly (i.e., at the end of each week) to ensure levels remained above 80 percent agreement. Due to scheduling conflicts (e.g., teacher absences, school holidays,

external observer availability), less than 20 percent of observations in Phase A included two external observers. However, when totaled across all phases and separately for Phases B and C, 20 percent or more of observations included two observers.

Interobserver agreement fell below 80 percent after the first week of data collection, so the primary investigator and external observer discussed their ratings and reviewed each construct on the DBR-CM in detail. The external observer demonstrated a clear understanding of why the disagreements between ratings occurred and was referred to the DBR-CM website for further training if needed. Additionally, the primary investigator encouraged the observer to reference the more detailed definitions of each item toward the bottom of the form when completing ratings after an observation.

Fidelity. To ensure high-quality delivery of study PD activities, implementation fidelity was monitored throughout treatment phases (i.e., Phases B and C). Fidelity was monitored across adherence (i.e., number of intervention components completed), quality (e.g., enthusiasm implemented), and duration (i.e., number of treatments completed; time to complete each treatment). The primary investigator completed a study specific measure of implementation fidelity daily during treatment phases (i.e., Phases B and C). Adherence measures varied slightly across phases, with Phase C (i.e., performance feedback with self-monitoring) including items addressing self-monitoring components (e.g., Teacher completed DBR-CM SR ratings). Though there are no clear guidelines regarding minimally acceptable levels of treatment fidelity, research has demonstrated the importance of a high level of adherence for interpreting study results (Sanetti et al.,

2021). Adherence was calculated weekly and remained above 90% for the duration of the study.

PD Activities (i.e., treatment/intervention)

***Training*.** After providing consent, participants met with the principal investigator to revisit the purpose and goals of the study, the general study timeline, and data collection procedures. Additionally, in this initial meeting, the principal investigator obtained copies of each teachers schedule and worked with them to identify blocks of core instructional time in which to conduct observations. Following the initial meeting participants entered the baseline phase (i.e., Phase A) and were observed by the principal investigator daily for 20 minutes to complete DBR-CM ER ratings. After at least five observations and stability (i.e., reasonable predictability) in one participant's DBR-CM ER Total Score was reached, that participant moved to Phase B (i.e., performance feedback PD activity).

Prior to commencing Phase B for participant 1 (i.e., the first participant to move to Phase B), the principal investigator briefly met with this participant to review procedures for this phase. During this meeting, the principal investigator informed teachers of the sequence of events for this portion of the study (see below for details). Generally, participants were informed that Phase B would include: 1) observation by the principal investigator, 2) formatting of observation results into a graphic format, 3) presentation of graphed results prior to the next observation period (i.e., before the next observed instructional block), 4) participant review of results, including descriptions of

items, 5) opportunity for participant questions, 6) performance feedback from principal investigator, and 7) repeat this process.

When transitioning from Phase B to Phase C, the principal investigator met with participants to cover procedures for this phase (see below for details) and to prepare them to complete DBR-CM SR ratings. Participating teachers were then emailed a link to the 40-minute, web-based DBR-CM training (https://dbr-cm.com/online-training) and instructions for how to complete the training reliability checks. Generally, participants were informed that Phase C would include: 1) completion of DBR-CM ER ratings by the principal investigator and completion of DBR-CM SR ratings by participants during the same observation period, 2) formatting of observations results (i.e., ER and SR scores) into a graphic format, 3) presentation of graphed results prior to the next observation period (i.e., before the next observed instructional block), 4) participant review of results, including descriptions of DBR-CM items, 5) opportunity for participant questions, 6) performance feedback from principal investigator, and 7) repeat this process. Phase C continued for a minimum of one school week (i.e., 5 feedback sessions) and/or until stability was established across DBR-CM ER Total Scores for Participant 1. Participants were asked to complete the online DBR-CM training within 48 hours to facilitate timely transitions to Phase C. However, completion times varied from one to eight days which delayed the transition to Phase C for some participants.

Baseline. The three initial observations that were conducted for screening purposes were included as part of baseline data collection. During the baseline phase, teachers were asked not to alter their usual teaching or CM practices. Observations

occurred during the core instructional periods specified by participants during the initial orientation meetings. Only data from the final baseline data observation (i.e., scores from the observation immediately prior to beginning Phase B) were shared with teachers. These scores were shared with teachers using the structured performance feedback procedures for Phase B (see below).

Performance feedback. Phase B continued for a minimum of one school week (i.e., 5 feedback sessions) and/or until stability was established across DBR-CM ER Total Scores for Participant 1. After stability in Participant 1's DBR-CM ER Total Scores was noted and at least 5 data points were collected, Teacher 1 moved into Phase C (i.e., performance feedback with self-monitoring). Transitions from Baseline Phase to Phase B and Phase B to Phase C followed noted transition rules for remaining participants. The first treatment (i.e., Phase B; performance feedback PD activity) began after the collection of at least five baseline data observations and when DBR-CM ER Total Scores appeared stable (i.e., limited variability, reasonably predictable trend). At the onset of Phase B, the DBR-CM item scores from the final baseline data observation were input into a graphic format (i.e., bar graph) and provided to participants with operational definitions of each item. The first performance feedback session occurred immediately prior to the first data collection point for Phase B. The principal investigator provided the teacher the graphic representation of their DBR-CM scores (see Appendix D). The principal investigator reminded the teacher that there are four DBR-CM items and briefly covered the operational definitions for each item. After a review of DBR-CM items, the principal investigator asked the teacher to review their scores from the previous

observation (i.e., the last Baseline Phase observation). The teacher was then given two minutes (as needed) to review their scores and reflect on their use of CM practices from the previous observation period. After the brief review period, the principal investigator solicited the participant's thoughts on the data (i.e., "What stands out for you on this graph?"). After allowing the participant to review their results, the principal investigator reflected and summarized the participants thoughts, if any. Next, the primary investigator solicited permission to provide verbal feedback regarding the participant's DBR-CM scores from the previous observation. The principal investigator noted at least one positive aspect of the data (e.g., "Your Praise score was your highest rating.") and one area for improvement (e.g., "It looks like there's room here to improve your Enthusiasm rating."). The principal investigator then inquired as to whether the participant had any questions about the data or study procedures. This brief, 5 to 10-minute meeting concluded with a reminder of the next scheduled observation time.

This process continued through the duration of Phase B. During each performance feedback session, data collected in the prior observation was presented to respective participants prior to the observation period for that day alongside their data from previous observation periods (if applicable). Performance feedback included both visually and verbally presented information. Teachers reviewed the graph showing their DBR-CM ER scores separated and color coded by construct (e.g., praise, communication, etc.), emailed in electronic PDF form. This procedure of emailing teachers a graph of their scores each day was modified, based on participant preference, to a google sheets link that was updated daily and displayed on the primary investigators' laptop during feedback

sessions. Teachers were given two minutes (as needed) to review their results and reflect on the previous observation period. Participants were given the opportunity to share their reflections and ask questions. The principal investigator then provided reflections, summaries, and brief unique feedback (when needed) (i.e., strengths and opportunities for improvement). Each performance feedback session concluded with reminders of the next session and brief positive verbal reinforcement for the participants' efforts to implement positive CM practices and participate in study activities.

Performance Feedback plus Self-Monitoring. The second treatment (i.e., Phase C; performance feedback with self-monitoring) began after the collection of at least one week (i.e., 10 feedback sessions) of Phase B data collection observations and when DBR-CM ER Total Scores appeared stable (i.e., limited variability, reasonably predictable trend). As noted previously, immediately preceding the onset of Phase C, participating teachers met with the principal investigator to prepare for DBR-CM SR completion (i.e., online training; overview of phase procedures). After this meeting and completing preparation activities (i.e., training and reliability check), participants moved to Phase C. Unlike Phase B, Phase C did not utilize DBR-CM scores from the last observation in the previous phase. Phase C began with a pre-observation meeting to remind the teacher they would be completing the DBR-CM SR at the conclusion of the observation period using the link provided to the Qualtrics formatted rating form. At the conclusion of the 20-minute observation, the external observer prompted the teacher to complete their DBR-CM SR ratings and provided the link to the form via text or email according to participants' preference. The primary investigator used caution when delivering this

prompt and waited for a natural pause in instruction so as not to interrupt or disrupt the flow of the lesson. Next, DBR-CM ER and SR scores were input into a graphic format (i.e., bar graph) and participants were provided with operational definitions of DBR-CM items. Formatting allowed for "side-by-side" comparison of ER and SR ratings with each set of ratings clearly labeled with self-report or external rater. The first performance feedback with self-monitoring session occurred prior to the next observation period. The principal investigator provided the teacher the graphic representation of their DBR-CM scores (see Appendix D) first via emailed PDF but later changed to a google sheets link according to participant preference. The principal investigator reminded the teacher of the four DBR-CM items and briefly covered their operational definitions. After a review of DBR-CM items, the principal investigator allowed the teacher to review their scores from the previous observation (i.e., the first Phase C observation), noting the ER and SR scores respectively. The teacher was then given two minutes, as needed, to review the scores and reflect on their use of CM practices from the previous observation period. After the brief review period, the principal investigator solicited the participant's thoughts on the data (i.e., "What stands out for you on this graph?"). After allowing the participant to contemplate their results, the principal investigator reflected and summarized the participants thoughts, if any. Next, with permission, the primary investigator provided verbal feedback regarding the participant's DBR-CM scores from the previous observation. The principal investigator noted at least one positive aspect of the data (e.g., "Your Praise score was your highest rating.") and one area for improvement (e.g., "It looks like there's room here to improve your Enthusiasm rating."). When possible, the

principal investigator highlighted agreement and/or divergence between ER and SR scores. The principal investigator offered the participant the opportunity to ask additional questions about the data or study procedures. This brief, 5–10-minute meeting concluded with a reminder of the next observation period and that participants would be rating their own classroom management during this period along with the external observer.

This process continued through the duration of Phase C. Each performance feedback session, the data collected the prior observation was presented to respective participants prior to the observation period for that day. Performance feedback included both visually and verbally presented information, with care taken to focus on noteworthy ER and SR ratings comparisons. Teachers review a graph showing their DBR-CM ER and SR scores separated and color coded by construct (e.g., praise, communication), emailed in electronic PDF form or via a google sheets link, based on participant preference. Teachers were given two minutes as needed to review results and reflect on the previous observation period. Participants were given the opportunity to share their reflections and ask questions. The principal investigator provided reflections, summaries, and brief unique feedback (when needed) (i.e., strengths and opportunities for improvement; agreement or disagreement in ER and SR ratings). Each performance feedback session concluded with a reminder for the next session and brief positive verbal reinforcement for the participants' efforts to implement positive CM practices and participate in study activities.

Results

Participant Demographics

Teachers were asked to complete a demographic survey prior to the start of data collection. The survey included information regarding highest level of education, grade taught, years of experience teaching, gender, race/ethnicity, and degree type. All four teachers included in the study identified as non-Hispanic Caucasian females. All teachers completed a bachelor's degree in general education as their highest degree. Number of years of teaching experience for teachers 1 through 4 was 3, 2, 4, and 4 years, respectively. Grade taught for teachers 1 through 4 was 3rd, 5th, 3rd, and Kindergarten, respectively. Teacher 5, who did not meet study inclusionary criteria, identified as a Hispanic Caucasian female. Teacher 5 reported having 15 years teaching experience, currently teaching Kindergarten, and highest degree held was a bachelor's degree in general education.

Fidelity

To ensure high quality delivery of study PD activities, implementation fidelity was monitored throughout treatment phases (i.e., Phases B and C). Fidelity was monitored across adherence (i.e., number of intervention components completed), quality (e.g., enthusiasm implemented), and duration (i.e., number of treatments completed; time to complete each treatment). The primary investigator completed a study specific measure of implementation fidelity daily for each participant during treatment phases (i.e., Phases B and C). This checklist was used for calculating implementation fidelity and was recorded using a google form. The measure included checks for adherence to the

PD procedures for both (a) delivery of performance feedback and self-monitoring activities and (b) duration of the PD activities (minimum 3 minutes to allow for review of scores and opportunity for questions).

Implementation fidelity, as measured by daily adherence checklists, was 100% in both Phase B and C indicating that study procedures were consistently implemented as written. Quality was measured using a five item, five-point Likert-type rating scale. This study specific quality measure contained items regarding graphic feedback clarity, verbal feedback clarity, helpfulness, tone, and timeliness of the feedback session. Teacher participants anonymously completed these ratings after each performance feedback session that occurred during treatment phases and scores were averaged. All teachers across both phases and indicated "strongly agree" for every item on the study specific measure of implementation quality resulting in an average score of 5. These consistently high scores indicate high quality implementation fidelity for treatment phases. The final measure of fidelity was duration which recorded the number of treatments completed and the time to complete each treatment. Results indicated 71 treatments completed with an average duration of 6 minutes per feedback session across both treatment phases. During Phase B, 30 treatments were completed and during Phase C, 41 treatments were completed. The average session duration in Phase B was approximately 7 minutes and the average session duration in Phase C was approximately 5 minutes.

Visual Analysis

In single-case design, visual analysis of the data is used to determine if there is evidence of a relationship between the independent (performance feedback interventions)

and dependent (teacher CM behaviors) variables and the strength of that relationship (Kratochwill et al., 2010). An effect is shown when the pattern of the data in one phase differs more than would be expected based on the pattern of the data that was observed in the previous phase (Kratochwill et al., 2010). It is also important that the effect is seen with some immediacy after the implementation of the intervention phase (Horner et al., 2005). This means there should be only a short delay between the manipulation of the independent variable (IV) and the change observed in the dependent variable (DV) (Horner et al., 2005). More specifically, there are four steps to visual analysis and six features that can be used to examine within and between phase data patterns (Kratochwill et al., 2010).

To briefly outline the four steps, (1) document a predictable pattern of data in the baseline phase; (2) assess within-phase patterns for each phase of the study; (3) compare the data from similar phases to determine if manipulating the independent variable had an effect on the data; (4) look for at least three demonstrations of an effect by integrating information from all phases of the study (Kratochwill et al., 2010). In addition to these four steps, there are six features of visual analysis. The six features of visual analysis (i.e., level, trend, variability, immediacy of the effect, overlap, and consistency of data patterns across similar phases) outlined by WWC help determine whether there was a causal relationship between the IV and the DV. Figure 1 displays Teachers graphed DBR-CM total scores across each phase of the study. Table 1 includes mean and standard deviation for each phase and for each teacher which were used to support visual analyses

Table 1

Mean and Standard Deviation of DBR-CM Total Score by Teacher and Phase

		Mean	SD	*n*
Teacher 1	Baseline	21.8	2.04	5
	Phase B	31.4	2.61	7
	Phase C	36.3	2.37	14
Teacher 2	Baseline	25.1	4.20	9
	Phase B	31.2	3.02	6
	Phase C	35.1	2.50	11
Teacher 3	Baseline	20.2	6.50	11
	Phase B	35.0	2.04	10
	Phase C	36.6	3.40	7
Teacher 4	Baseline	18.5	4.40	17
	Phase B	31.7	2.37	7
	Phase C	34.9	2.66	8
Total	Baseline	20.8	5.5	42
	Phase B	32.8	3.7	30
	Phase C	35.7	2.8	41

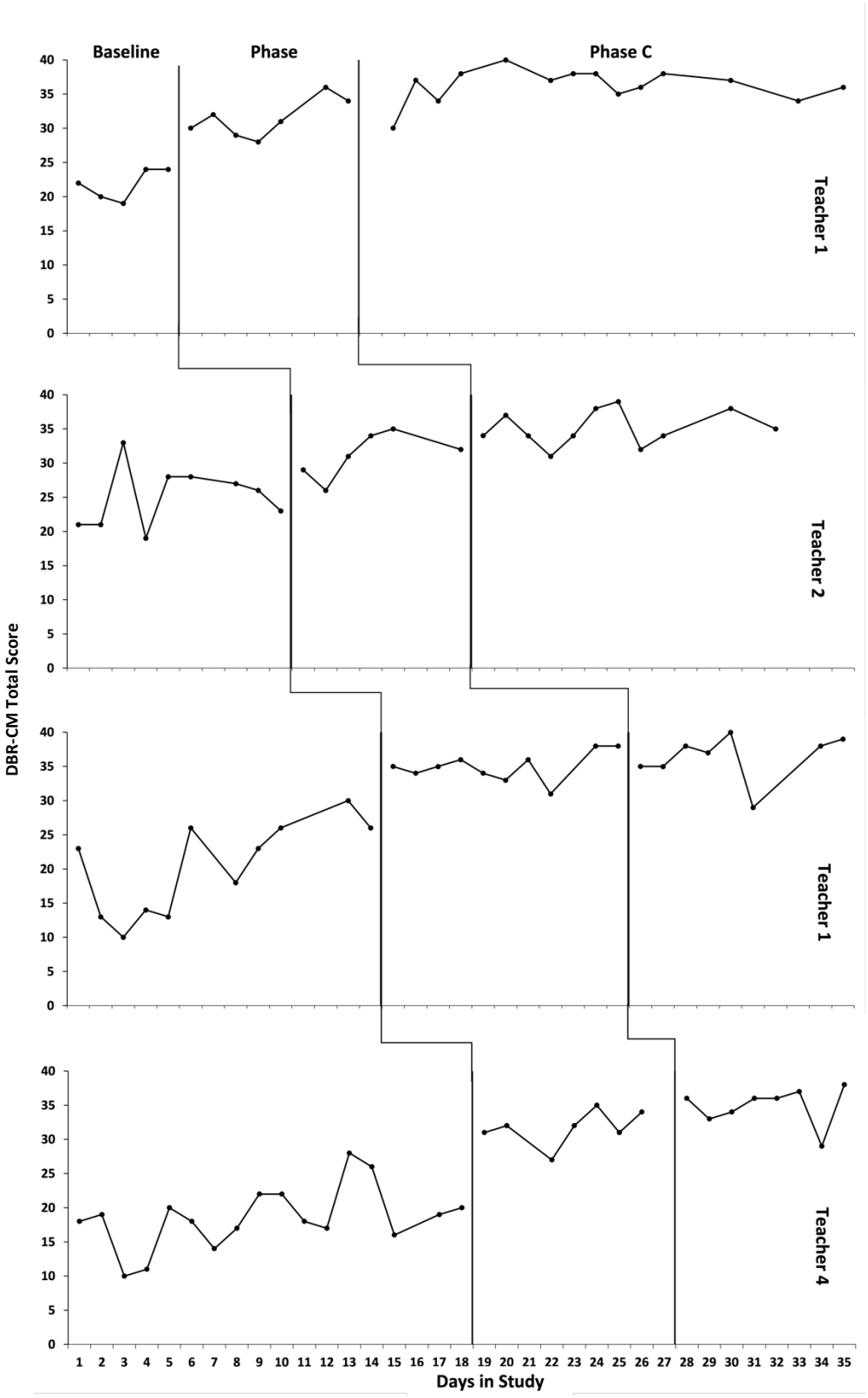
Baseline
Phase
Phase C
Teacher 1
Teacher 2
Teacher 1
Teacher 4
DBR-CM Total Score
Days in Study

Figure 1

Visual Presentation of Participant Data: DBR-CM Total Score

Note. The figure above shows graphed data for the combined multiple-baseline ABC design implemented in the present study. Data points in each phase are connected with a continuous line despite gaps in observational data for a given teacher (e.g., absence).

Teacher 1

Visual analysis of DBR-CM ER total score for Teacher 1 indicates a generally positive response to the first intervention (Phase B) and a neutral to slightly positive response to the second intervention (Phase C). When considering a change in level, or the apparent average score of data within a condition (Riley-Tillman & Burns, 2009), there appears to be an increase in level from baseline to Phase B and another less pronounced increase from Phase B to Phase C. To support this visual analysis of level based on the graphed data, level was also assessed by calculating the mean score in each phase for each teacher and comparing scores. In looking at the trend of the data, or the direction the data is going in on the graphs, there appears to be no clear trend in the data in baseline phase, an increasing trend in Phase B, and a slightly increasing trend in Phase C. Visual analysis of variability, or the range or standard deviation of data around the trend line (Riley-Tillman & Burns, 2009), between baseline and Phase B shows slightly less variability in Phase B. Graphed data in Phase C appears to have the least amount of variability which indicates a higher degree of control of the intervention. Another important feature of visual analysis is immediacy of the effects of the intervention which indicate how quickly behaviors changed from baseline to intervention. When comparing the last three data points in baseline to the first three data points in Phase B there appears

to be an immediate effect of the first intervention. There did not appear to be an immediate effect of the second intervention (Phase C) when compared to the first intervention (Phase B). Finally, analyzing overlap of the data shows no overlap in data points from baseline to Phase B and about half of the data in Phase C overlap with Phase B. The less overlap that exists, the more likely it is that the intervention was effective in changing behavior from baseline to intervention. Overall, visual analysis of data for teacher 1 suggests there were intervention effects.

Teacher 2

Visual analysis of DBR-CM ER total score for Teacher 2 indicates a neutral to slightly positive response to the first intervention (Phase B) and the second intervention (Phase C). Overall, the level of data in Phase B appears to be higher than the level of data in baseline phase. When comparing level in Phase C to Phase B, there appears to be a slight increase in level. These conclusions are supported by calculating and comparing the mean score in each phase. In looking at the trend of the data there seems to be a slightly positive trend in the data in the Phases B and C, while no clear trend is seen in the baseline phase. Visual analysis of variability between baseline phase and Phase B indicates less variability in Phase B. Variability in Phase B and Phase C appear to be similar. Teacher 2 also shows immediacy of the effects of the intervention between baseline and Phase B but not between Phase B and Phase C. Finally, there appears to be a large amount of overlap in data points from the baseline phase to Phase B which would indicate that the intervention did not have an effect. However, there appears to be much

less overlap between Phase B and Phase C. Overall, visual analysis of data for teacher 2 is inconclusive but may suggest slight intervention effects.

Teacher 3

Visual analysis for Teacher 3 shows a trend in the baseline phase which makes it more difficult to draw conclusions about the effects of the interventions so results should be interpreted with caution. In comparing level of the data in Phase B to level in baseline, Phase B shows a clear increase in level. There is no visible change in level from Phase B to Phase C. Visual analysis of trend for Teacher 3 shows a positive trend in the baseline phase and no clear trend in Phase B or Phase C which limits the interpretation of intervention effects. While a positive trend can be observed in baseline phase, there is also considerable variability in the data in this phase while Phase B graphs show stability in the data with little variability. Comparison of variability in Phase B and C indicates more variability in Phase C than in Phase B. Teacher 3 also shows immediacy of the effects of the intervention between baseline and Phase B but not between Phase B and Phase C. When looking for overlap in the data between baseline phase and Phase B, none can be observed. There appears to be lot of overlap in the data between Phase B and C, however. Overall, visual analysis of data for teacher 3 indicates some positive effects of the first intervention when compared to baseline, but results of this visual analysis have limited interpretability due to the positive trend in the data that can be observed in baseline phase.

Teacher 4

Results of visual analysis for the final participant, Teacher 4, suggest that the first intervention had a positive effect on DBR-CM total score. The second intervention (Phase C) did little to add to this positive effect when compared to data in the first intervention (Phase B). A sizable increase in level can be observed from baseline phase to Phase B and a less dramatic though still noticeable increase can be observed from Phase B to Phase C. Again, these observations are supported by calculating and comparing the mean score in each phase. In looking at the trend of the data in each phase, there seems to be a slightly positive trend in the data in the Phases B and C, while no clear trend is seen in the baseline phase. Visual analysis of variability between baseline phase and Phase B indicates less variability in Phase B. There is also little variability in Phase C data when compared to Phase B data. Teacher 4 also shows immediacy of the effects of the intervention between baseline and Phase B, but no notable change occurs from Phase B to Phase C. Finally, there appears to be very little overlap in the data between baseline and Phase B as well as between Phase B and Phase C which would indicate that the intervention did have an effect.

Empirical analyses

To support visual analysis of data collected, non-overlap statistics and effect size were calculated for DBR-CM ER Total Score. This was done following procedures outlined in Riley-Tillman et al. (2020) for Percent of Non-overlapping Data (PND), Percent of All Non-overlapping Data (PAND), and Tau-U.

PND

To calculate PND, a straight horizontal line was drawn from the highest positive baseline data point (i.e., highest DBR-CM ER total score) through the intervention data points in Phase B. Then, the number of intervention data points above this line was divided by the total number of intervention data points to get the percent of nonoverlapping data or PND. In the present study, PND was calculated between Phase B (intervention 1) and baseline, as well as between Phase C (intervention 2) and Phase B (intervention 1). Because all participants received intervention 1 prior to beginning intervention 2, no conclusions could be drawn from calculating PND between Phase C and baseline. 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 89% indicate effective interventions; PND scores that fall between 51% and 69% indicate interventions with questionable effect; and a PND score of 50% or lower indicate an intervention was ineffective. PND between baseline phase and Phase B was calculated for teachers 1 through 4 as follows 100%, 33%, 100%, 86%. PND between Phase B and Phase C for teachers 1 through 4 was calculated as follows 57%, 36%, 29%, and 63%. All PND calculations for all teachers and phases can be found in Table 2.

PAND

PAND follows a different and slightly more complicated procedure that uses all the data rather than just the most extreme point and it considers the data from each participant together rather than averaging individual scores. PAND equals 100 percent

minus the total percentage of intervention data points that overlapped with baseline data points across all participants. PAND was calculated between baseline phase and Phase B which yielded a PAND of 83%, and between Phase B and Phase C which resulted in a PAND of 48%. PAND scores of 50% and below reflect chance level overlap between phases and are interpreted as having no intervention effect. Scores between 70% and 89% indicate effective interventions.

Tau-U

Finally, effect size was calculated using Tau-U which can be used as a single estimate of effect of an intervention that controls for confounding baseline trend if needed. Tau-U is particularly well-suited for two-condition studies and small datasets that do not conform to typical parametric assumptions like the present study (Riley-Tillman et al., 2020; Sims, 2016). Tau-U statistics were calculated using a web-based calculator developed by Parker and colleagues (2011) and can be found in Table 2. First, contrasts were run to determine if there were any trends in the baseline phase or Phase B (which served as a baseline for the second intervention) of each teacher. Results indicated an increasing trend in baseline for Teacher 3, but no other participants showed a trend in either of the two 'baseline' phases in the study. Due to the significant trend in baseline for teacher 3, Tau-$U_{adj\ was}$ calculated for this participant to give an estimate of effect while controlling for the positive trend in baseline phase. For all other phases and teachers, Tau-U was calculated without controlling for a trend in baseline. An overall estimate of intervention effect was not able to be calculated because the same Tau-U was not used for all cases (Fingerhut et al., 2021).

Table 2

Advanced Empirical Analyses of Intervention Effects on Classroom Management Total Score: Baseline vs Phase B and Phase B vs Phase C

	Baseline vs Phase B			*Phase B vs Phase C*		
	PND	Tau U	PAND	PND	Tau U	PAND
Teacher 1	100%	1.0*	-	57%	.81*	-
Teacher 2	33%	.72*	-	36%	.62*	-
Teacher 3	100%	.73*	-	29%	.46	-
Teacher 4	86%	.98*	-	63%	.66*	-
Total	-	-	83%	-	-	48%

Note. PND <50% indicates no intervention effect. Tau-u <.65 indicates weak or small effect.

[a]Tau-U_{Adj} was used for Teacher 3 in Baseline vs. Phase B calculations.

* $p < 0.05$ level.

Interobserver Agreement

To ensure reliability of study data and to meet WWC standards, 20% of observations across all phases were scheduled to include two external observers who would complete concurrent observations and DBR-CM ER ratings. However, due to external observer availability, teacher absences, and school closures, the percent of observations where IOA was able to be calculated fell below 20% in baseline phase across all participants. The number of observations with two external raters also fell below 20% for Teachers 1, 2, and 4 when totaled across all phases by participant. For the DBR-CM ER, reliability was calculated by comparing ratings in each category for a specific observation period and calculating the percentage of categories that were rated within one point of each other. Percent of observations where IOA was calculated as well as percent agreement for each teacher and phase are displayed in Table 3. IOA ranged

from 83% to 95% in each condition for each participant. IOA was calculated for 20% of all observations across all participants and conditions with a total percent agreement of 91%. IOA was also calculated between DBR-CM ER ratings and DBR-CM SR ratings during Phase C. Agreement was calculated for 100% of observations in Phase C because participants were required to rate their own classroom management practices after each observation in this phase. Levels of IOA for Teachers 1 through 4 were 86%, 43%, 79%, and 88%.

Table 3

Interobserver Agreement for DBR-CM Item Scores

Phase	% of Observations	% Agreement
Baseline	14	83
B	23	93
C	22	94
Participant	**% of Observations**	**% Agreement**
Teacher 1	15	88
Teacher 2	19	95
Teacher 3	24	93
Teacher 4	19	88
Total	20	91

Social Validity

To evaluate teacher perceptions of acceptability, understanding, feasibility, system climate, and system support of an intervention, teachers were asked to fill out the User Rating Profile, Intervention Revised (URP-IR) for Phase B and Phase C (Briesch et al., 2013). Results of teacher ratings on the URP-IR for both Phase B and Phase C yielded scores that indicate these interventions are moderately socially acceptable. The

results of teacher ratings on the URP-IR for Phase B are reported in Table 4 and Rating for Phase C are reported in Table 5. Acceptability scores for Phase B (performance feedback only) and for Phase C (performance feedback plus self-monitoring) ranged from 4 to 6 and resulted in a mean score of 5 (1 = *Strongly Disagree*, 2 = *Disagree*, 3 = *Slightly Disagree*, 4 = *Slightly Agree*, 5 = *Agree*, and 6 = *Strongly Agree*). The acceptability subscale measures a rater's enthusiasm about implementing an intervention. Understanding scores for Phase B and for Phase C ranged from 5 to 6 and resulted in a mean score of 6. 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 for Phase B and Phase C interventions were a 5, with individual scores ranging from 5 to 6. This subscale measures the rater's confidence in his or her ability to implement the intervention with integrity given existing demands. Rater responses yielded Systems Support subscale scores that ranged between 1 and 4 for Phase B and between 1 and 3 for Phase C. This resulted in a mean score of 3 across all raters for Phase B and a mean score of 2 for Phase C. 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. Finally, teacher ratings on the System Climate subscale ranged from 5 to 6 for both intervention phases and resulted in a mean score of 5 across all raters for both interventions. The Systems Climate subscale measures the extent to which a respondent believes that the intervention would be a good fit for their school system.

Table 4

Performance Feedback Intervention
User Rating Profile-Intervention Revised, Average Item Score

	Acceptability	Understanding	Feasibility	System Climate	System Support	Total
Rater 1	5	6	5	6	5	5
Rater 2	6	6	6	6	6	6
Rater 3	5	5	5	5	3	5
Rater 4	4	6	5	5	3	4
ALL	5	6	5	5	4	5

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

Table 5

Performance Feedback and Self-Monitoring Intervention
User Rating Profile-Intervention Revised, Mean Item Score

	Acceptability	Understanding	Feasibility	System Climate	System Support	Total
Rater 1	6	6	5	6	5	6
Rater 2	6	6	6	6	6	6
Rater 3	5	5	5	5	4	5
Rater 4	4	6	5	5	4	5
ALL	5	6	5	5	5	5

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

Discussion

The purpose of this study was to evaluate the effects of a performance feedback intervention and a performance feedback plus self-monitoring intervention on teacher's classroom management behavior. Classroom management practices for four teacher participants were assessed (almost) daily using the DBR-CM. An external observer (i.e., the primary investigator) completed the DBR-CM ER after each observation and used those scores to provide teacher with both visually and verbally presented feedback prior to the start of the next observation session during Phase B. Teachers also engaged in a combined performance feedback with self-monitoring PD activity using the DBR-CM SR during the second portion of the study. The two interventions were staggered across four subjects over 35 school days. Results will be discussed in relation to the four primary research hypotheses guiding this study.

Research Hypotheses

Performance Feedback Intervention

Research Hypothesis 1: Performance feedback-based PD will increase teacher use of evidence-based CM practices. Data collection across all four subjects for Phase B of the study (performance feedback only) demonstrated the minimum three replications of an effect that are required to establish intervention effectiveness (Kratochwill et al., 2012). Both visual and empirical analysis of the data for Teachers 1, 3, and 4 demonstrates a moderately strong positive effect of the intervention on teachers' classroom management practices as measured by DBR-CM ER total score. All three teachers showed a clear increase in level from baseline to Phase B and a decrease in

variability between baseline and Phase B. Teachers 1, 3 and 4 also demonstrated immediacy of the effects of the intervention and little to no overlap between phases. Trend was a major concern that emerged during visual analysis of the data. While all three teachers showed a positive trend during intervention phase, Teacher 3 also showed a slightly positive trend in the baseline phase which limits the interpretability of visual analysis of intervention effects. However, empirical analysis of the data indicated that a positive effect of the intervention was still present after using baseline correction procedures. PND was calculated for each teacher as part of empirical analysis and yielded scores of 100%, 100%, and 86% for teachers 1, 3, and 4 respectively. PND scores at or above 90% were considered *very effective*, between 70% and 90% were considered *effective*, between 50% and 70% were considered *questionable*, and below 50% were considered to be *ineffective* or to have no effect (Scruggs & Mastropieri, 2001). However, the PND score for Teacher 3 has limited interpretability due to the trend in data seen in their baseline phase. Empirical analyses also included calculation of Tau-U for each teacher which provides an option to control for a trend in the baseline. Tau-U was calculated for Teachers 1, 2, and 4, but Tau-U_{adj} was used for Teacher 3 to correct for the trend found in the baseline phase. Tau-U scores for Teachers 1, 3, and 4 were 1, .73, and .98 respectively. Guidelines for interpreting Tau-U statistics by Parker and Vannest (2012) state that Tau-U scores below .65 indicate weak or small intervention effects, scores between .66 and .92 indicate medium to high intervention effects, and scores above .93 indicate large or strong intervention effects. Based on these guidelines, Teacher 3 demonstrated medium to high intervention effects even when correcting for the

trend in baseline phase data.

Visual analysis of teacher 2 indicates possible intervention effects but further analysis is required to make a determination due to the significant overlap in data between baseline and Phase B. Visual analysis for Teacher 2 indicated an increase in level from baseline to Phase B, a slightly positive trend in Phase B with no trend in baseline, decreased variability in Phase B, and immediacy of effects of the performance feedback intervention. However, a large amount of overlap was seen in the data between baseline and Phase B. Empirical analysis for Teacher 2 yielded a PND score of 33% and a significant Tau-U score of .72. PND is calculated using the highest baseline data point which, for Teacher 2, appeared to be an outlier compared to other baseline data points. Due to this one unusually high score in baseline phase, interpretation of PND score would indicate that the intervention was ineffective. In this case, Tau-U, which indicates medium to high effects, may provide a more robust estimate of the effectiveness of the intervention for Teacher 2.

To conclude, research hypothesis 1 is supported by both visual and empirical analysis of the data for teachers 1, 3, and 4. Teacher 2 also showed some intervention effects when the results of all visual and empirical analysis are considered together. However, results within each type of analysis are inconsistent. These findings indicate that teachers showed improvements in their classroom management practices, as measured by the DBR-CM, when provided with brief verbal and visual performance feedback before each observation period.

Performance Feedback & Self-Monitoring Intervention

Research Hypothesis 2: Combined performance feedback with self-monitoring PD will increase teacher use of evidence-based CM practices. Phase C of the present study combined performance feedback with a self-monitoring intervention where teachers rated their own classroom management behavior. Similar to procedures for Phase B, teachers were provided with brief verbal and visual (i.e., graphed) feedback on their scores but with the addition of seeing their own ratings included. Due to a previous similar study that employed a performance feedback intervention using the DBR-CM with little success (Sims, 2016), Phase C was added to the present study as an additional means for improving teacher classroom management behavior. In this instance, Phase B acted as the "baseline" for Phase C and scores were compared between phases to determine if the second intervention had a noticeable effect on teachers' classroom management after they had already received the first intervention. One concern with comparing an intervention phase to another intervention phase was the potential presence of a trend in the Phase B (i.e., the acting baseline for Phase C) which would give the results limited interpretability. However, trend analysis did not reveal any significant trends in Phase B. Data collection across all four subjects in Phase C revealed mixed results and did not meet the required three demonstrations of an effect of an intervention (Kratochwill et al., 2012). Visual analysis of the data for each teacher demonstrated a slight increase in level for teachers 1, 2, and 4 while no visible increase was seen for teacher 3. A slightly increasing trend in Phase C and Phase B were noted for teachers 1, 2, and 4. As previously discussed, the existence of a desirable (i.e., positive) trend in the baseline

phase (i.e., Phase B) limits the conclusions that can be drawn about the effects of an intervention. However, due to the nature of this study, trend analysis was computed in anticipation of this possibility which revealed no significant trends were present in Phase B for any participant. In comparing variability between the two phases, Phase C appeared to have very little variability in the data for teacher 1, 2, and 4. Teacher 3 showed slightly more variability in the data in Phase C than in Phase B which indicates that the second intervention did not result in increased control of the data for that teacher. Only Teacher 1 showed immediacy of the effects of the intervention with an increase in scores from the end of Phase B to the beginning of Phase C. Something to consider when analyzing this finding is that Teacher 1 received the first intervention for the shortest period of time, compared to the other three participants, before moving to Phase C intervention. Thus, Teachers 2 through 4 had more time in the first intervention phase for their scores to improve before they started Phase C interventions which could explain the lack of immediacy of effects for these teachers. Finally, overlap was visually analyzed between Phases B and C for each teacher which found a lot of overlap for teachers 2 and 3 and little overlap for Teacher 4. About half of the data points for teacher one appeared to overlap between Phases B and C. Effects of the intervention were observed through visual analysis of level, trend and variability for teachers 1, 2, and 4. However, overlap and immediacy of effects were not consistent in demonstrating an effect of the intervention for teachers 1, 2, and 4. For Teacher 3, none of the features of visual analysis indicated that the intervention had an impact on their classroom management behavior.

As with research hypothesis 1, PND and Tau-U were calculated for each teacher to compare Phases B and C and test the second hypothesis. PND between the two phases yielded scores of 57%, 36%, 29%, and 63% for teachers 1 through 4, respectively. PND for teachers 1 and 4 fell between 50% and 70% which indicates the intervention had questionable effects. For teachers 2 and 3, PND scores fell below 50% which indicates that the intervention was ineffective (Scruggs & Mastropieri, 2001). Considering that significant improvements in classroom management practices were already observed for most teachers from baseline to Phase B, the limited effects of the second intervention are not unexpected. However, further empirical analysis of intervention effects in Phase C based on Tau-U scores provide a different slightly different conclusion. Baseline corrected Tau-U was not necessary in this analysis because, despite the slight upward trend observed through visual analysis of the graphs, there were no significant trends in Phase B for any of the four participants. Tau-U scores for Teachers 1 through 4 were .81, .62, .46, and .66 respectively but Teacher 3's score was not significant at the $p < 0.05$ level. Guidelines for interpreting Tau-U statistics state that Tau-U scores below .65 indicate weak or small intervention effects, which would apply to Teacher 2. Teachers 1 and 4 have Tau-U scores that fall between .66 and .92 which indicate medium to high intervention effects. The score for teacher 4 is at the very bottom of this range indicating medium intervention effects while the score for teacher 1 is at the mid to upper end of these boundaries indicating higher intervention effects. Teacher 3 yielded an insignificant Tau-U score and did not contribute to the determination of intervention effects. Visual analysis was consistent with the findings of

empirical analysis for teacher 3, but only partially consistent with conclusions drawn from visual analysis for Teachers 1, 2, and 4 (how to make this sentence less confusing?).

To conclude, research hypothesis 2 was not true because the effect (i.e., increase in evidence-based classroom management practices) was not replicated for at least 3 participants which is the minimum number required to establish intervention effects in single case design. Slight improvements were observed for individual teachers but results within each type of analysis are inconsistent. A notable improvement in performance was observed at the onset of the first intervention (Phase B) which left less room for improvement at the onset of the second intervention (Phase C). All four teachers' average classroom management total scores from Phase B would have disqualified them from the study based on the initial screening criteria used for selecting participants (i.e., average DBR-CM total scores between 12 and 28). These inclusion criteria were developed based on the goal of the study which was to test the effects of a Tier 2 level professional development activity. Teachers who's average DBR-CM scores are above 28 would not be targeted for Tier 2 professional development because they have already demonstrated a certain level of mastery of classroom management practices and because they would have less opportunity to demonstrate significant improvements through intervention. Thus, it is unsurprising that significant improvement in classroom management practices was not observed from Phase B to Phase C considering the already increased scores. Based on the results of both visual and empirical analysis, hypothesis 2 was not supported because there were not the minimum three demonstrations of an effect of the intervention. However, findings indicate that the performance feedback plus self-

monitoring intervention did improve classroom management practices for some teachers to an extent.

Acceptability: Performance feedback

Research Hypothesis 3: Teachers will rate DBR-CM performance feedback as an acceptable PD practice. Hypothesis 3 relates to the social validity of the study and was evaluated using teachers URP-IR ratings of performance feedback. The URP-IR provides insight into participants views of the acceptability, feasibility, and understanding of an intervention. This rating scale also asks participants to rate their perception of system support required for an intervention and the level of "fit" of an intervention within a particular school system. Teachers generally rated the performance feedback intervention as an acceptable PD practice with scores ranging from 4 to 6 and 5 being the average score. Three out of four teachers found the intervention to be at least moderately acceptable with only one teacher indicating slightly acceptable. Teacher ratings of understanding indicated that teachers were highly confident in implementing the intervention and were able to easily comprehend the procedures required to do so. All but one participant (who scored a 5) scored a 6 on the understanding subscale with the average being 6 across all teachers. Based on their ratings, teachers found the performance feedback intervention to be feasible given their existing demands and the demands of the intervention. Feasibility scores ranged from very feasible (6) to feasible (5) with the average rating being 5. One teacher had a feasibility score of 6 (very feasible) while the remaining three participants had scores of 5 (feasible). System climate was rated similarly with the average score being 5. Half of participants indicated that this

intervention would be a good fit within their school system (5) and the other half rated the intervention as being a *very* good fit (6) for their school. These ratings indicate that the values and principles of this school likely align with the goals of the PD activity that was implemented. Ratings could also indicate that there is a perceived need for this type of PD practice at the school. Considering the emphasis on positive behavior interventions and supports (PBIS) that is promoted within the district where this study took place, these findings are not surprising. The final subscale is System Support which is meant to reflect participants beliefs that the intervention could not be implemented effectively without other internal supports from their school. For this subscale, lower ratings are more desirable because they reflect a greater ability to independently implement the intervention. On all other subscales, higher ratings are more desirable so to allow for data to be aggregated across all factors and for ease of interpretability, items from this subscale were reversed scored. The average score for this subscale was 4, indicating that participants were only slightly in agreement that this intervention could be implemented without additional system support. The average rating across all subscales and participants was a 5 which indicates that participants generally found this to be an acceptable PD activity. These ratings provide support for the third research hypothesis. Qualitative feedback from participants also indicated satisfaction with intervention activities and their outcome.

Acceptability: Performance Feedback & Self-Monitoring

Research Hypothesis 4: Teachers will rate the use of DBR-CM performance feedback and self-monitoring as an acceptable PD practice. As with Hypothesis 3, the

use of DBR-CM PF and SM as an acceptable PD practice was measured using teacher ratings across five subscales of the URP-IR. Participant ratings of performance feedback and self-monitoring on the URP-IR were nearly identical to ratings of the performance feedback only intervention. Acceptability, feasibility, system climate, and understanding subscales yielded the same average ratings across both PD activities (PF and PF +SM). Teachers rated both PD activities to be acceptable, feasible, and consistent with their system climate resulting in an average score of 5 on these subscales. Understanding was rated the most highly (average score of 6) indicating that teachers had a strong understanding of each of the two PD activities they participated in and were confident in their ability to implement them. System support was the only subscale that was rated differently by teachers across the two PD activities. According to their ratings, teachers were more confident about implementing the performance feedback plus self-monitoring intervention independently (i.e., without system support) than they were the performance feedback only intervention. The average rating for PF only was 4 (i.e., slight agreement) while the average rating for PF + SM was 5 (i.e., agreement). One possible explanation for this difference is that most teachers spent more time in the second intervention phase (PF + SM) so their familiarity with the intervention could have impacted their belief that it could implemented without system support.

Limitations

There are several notable limitations within this study. Frist, participant demographics reveal that the teachers included in this study were a very homogenous sample. While the similarity in the subjects may decrease potential moderating factors, it

also limits potential generalizability of findings. Next, a significant limitation of this study is the low percentage of observations with two raters present. For a study to meet WWC standards for single case design research, interobserver agreement must be calculated for a minimum of 20% of observations in each phase for each participant. Interobserver agreement helps ensure accuracy and reliability of data collection observations. The limited availability of the external rater along with there being only one external rater involved in data collection (besides the primary investigator), meant that IOA was calculated for less than 20% of observations for some phases and participants. Though the total percentage of observations with two raters across all observations for all participants in all phases was 20%, percentages for individual phases and participants fell below the minimum 20% required. The percentage of observations where IOA was calculated as well as level of agreement for each phase and for each teacher are shown in Table 3. Only 14% of all observations in the baseline phase included two raters which reduces the reliability of the data in this phase thus limiting the reliability of conclusions drawn from comparing baseline data to data in other phases. Additionally, percent agreement in this phase was 83% which was also lower than IOA calculated for phases B and C. As previously discussed, interobserver agreement was low at the start of data collection with the additional rater which prompted further training and discussion of DBR-CM scores to ameliorate disagreements. IOA greatly improved after these initial disagreements were addressed, but most disagreements occurred during baseline phase which affected the total percent agreement in this phase. When looking at the number of observations with two raters across teachers, only one teacher (teacher 3) met and

exceeded the minimum 20% set by WWC. Observations involving two raters for Teachers 1, 2, and 4 were 15%, 19%, and 19%, respectively which is a limitation of the present study. When considering the implications of these results, it is important to note that intervention effects were observed for two of the teachers that did not meet the minimum requirements for IOA observations. Thus, the absence of an adequate number of observations where IOA was calculated is a significant limitation of the present study. Another limitation can be identified within the design of the study. Direct effects of the Phase C (PF + SM) on teachers' classroom management practices could not be observed because each participant received the first intervention in Phase B (PF only) before starting Phase C (PF + SM). Instead, Phase C observational data was compared to Phase B, but this did not allow for any conclusions to be drawn about the utility of Phase C as an effective PD activity for teachers with moderate deficits in classroom management. The number of participants included in the present study could also be considered a limitation. Before beginning recruitment and data collection, the goal was to retain five teacher participants after screening all study volunteers. However, only five total participants volunteered to participate in the study and after screening teachers for inclusion criteria, only four participants were retained. Inclusion of one more participant would have provided an additional opportunity to observe the effects of the PD activities employed in this study. Finally, a positive trend was observed in the Baseline phase for Teacher 3 which limits the conclusions that can be drawn about their performance in subsequent phases of the intervention. Teacher 1 and Teacher 3 were both part of the same grade level team and their classrooms were in close proximity to one another at the

school. Because Teacher 1 was the first participant to begin receiving performance feedback, it is possible that discussion surrounding intervention activities took place between Teachers 1 and 3 without the primary investigators' knowledge. This discussion of study activities between participants is one explanation for the upward trend in Teacher 3's data which began shortly after Teacher 1 entered Phase B. Regardless of the reason for the baseline trend observed in Teacher 3's scores, that, along with their potential exposure to PD activities during baseline phase are both limitations of the present study.

Future directions

The present study sought to address several limitations noted in a previous study by Sims (2016) which used the DBR-CM to provide performance feedback with little success. Results of the present study indicate performance feedback using the DBR-CM can be an effective in producing increases in teacher CM behavior. However, the present study had its own set of limitations that could be addressed through replication of this research with some adjustments. The present study demonstrated some success in implementing the performance feedback only intervention using the DBR-CM before the second intervention was implemented making it difficult to draw conclusions about the effectiveness of the second intervention. Future applications of this research could look exclusively at the utility of the performance feedback plus self-monitoring intervention as a way to improve teachers classroom management practices. Another possible extension of this research is to examine the effectiveness of a self-monitoring only intervention using the DBR-CM. Previous research has demonstrated success using different self-

monitoring interventions to improve some aspect of teachers' performance but there is no available research supporting the use of the DBR-CM as a self-monitoring intervention to improve classroom management practices. Self-monitoring is a professional development practice that is generally effective, minimally invasive, and not resource intensive making this a valuable target for future research. While the present study was able to demonstrate an effect of the performance feedback intervention through visual and empirical analyses, the percentage of observations that included two raters was below the minimum amount required to meet WWC standards. Future replications of this study should aim to have a second rater available for at least 30% of observations to allow for unexpected schedule changes or absences without falling below the minimum of 20%. Adequate levels and amounts of IOA are essential for ensuring the accuracy and reliability of the data which in turn helps ensure the reliability of the conclusions being drawn from that data.

Conclusion

The present study examined a performance feedback intervention and a performance feedback plus self-monitoring intervention targeting teacher classroom management behavior. This intervention was based on previous research by Sims (2016) that demonstrated little success in improving teacher's classroom management behavior through a performance feedback intervention using the DBR-CM. One goal of both prior research and the present study was to develop a less time and resource intensive 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. Classroom management deficits were defined as an average DBR-CM total score between 12 and 28 and measured using three initial screening observations. Participants received brief verbal and visual performance feedback prior to each observation during the first phase of the study (Phase B). DBR-CM scores were presented to participants in a bar graph and briefly discussed before the start of the next observation. The second phase (Phase C) of the study involved both performance feedback and self-monitoring activities using the DBR-CM. In Phase C, participants rated their own classroom management using the DBR-CM and these scores were graphed and discussed alongside external observer scores prior to the next observation.

Data collected as part of this ABC multiple-baseline single case design study were analyzed in a manner consistent with prevailing guidelines and common practice (Kratochwill et al., 2010; Riley-Tillman & Burns, 2010). Visual and empirical analyses of results for Phase B indicated an improvement in overall teacher classroom management behavior in three of four teachers. However, standards set by WWC for the percentage of observations including two raters were not met during baseline Phase. Due to the high level of agreement among raters when IOA was calculated, it is believed that additional ratings would have yielded similar results.

Results of visual and empirical analyses for Phase C were mixed, indicating improvements for some teachers but no clear demonstration of intervention effects were seen when compared to data in the previous phase. These findings indicate that a brief

visual and verbal performance feedback intervention targeting classroom management was able to document the required three replications of intervention effects. While some Teachers did show marked improvement from Phase B to Phase C, results were generally inconsistent and did not establish three demonstrations of an effect of the second intervention. Due to the apparent increase in classroom management practices from baseline to Phase B, there was less opportunity for participants to demonstrate a noteworthy improvement in scores from Phase B to Phase C.

The present study was meant to build upon previous research done by Sims (2016) while addressing several limitations of that study. Modifications made to the study procedures used by Sims (2016) included: providing performance feedback prior to the start of the next observation rather than following the previous observation; the addition of a more formal training component for teachers which involved a discussion of DBR-CM items and scores prior to the first observation in Phase B and an online training and reliability check prior to Phase C; modifying performance feedback to include brief verbal feedback; presentation of visual feedback in bar graphs rather than line graphs for ease of interpretation; and careful consideration of both timing of the study as a whole (avoiding extended school breaks) and timing of individual observation periods (avoiding observations during non-instructional periods). One final, though unintentional, modification to Sims (2016) study procedures was including fewer participants which may have limited frustration at latency of the intervention for later participants.

Though this study was able to establish an intervention effect of performance feedback on classroom management using the DBR-CM it did not meet WWC standards

for percent of observations with two raters. Several other limitations and suggestions for future research are noted. The present study provides a compelling argument for both use of the DBR-CM as an effective Tier II performance feedback intervention and for future research that meets single-case design efficacy standards.

References

Alter, P., & Haydon, T. (2017). Characteristics of effective classroom rules: A review of the literature. *Teacher Education and Special Education*, *40*(2), 114-127.

Back, L. T., Polk, E., Keys, C. B., & McMahon, S. D. (2016). Classroom management, school staff relations, school climate, and academic achievement: Testing a model with urban high schools. Learning Environments Research, 19, 397.

Bayar, A. (2014). The Components of Effective Professional Development Activities in Terms of Teachers' Perspective. *International Online Journal of Educational Sciences*, *6*(2), 319-327

Becker, W. C., Madsen Jr, C. H., Arnold, C. R., & Thomas, D. R. (1967). The contingent use of teacher attention and praise in reducing classroom behavior problems. *The Journal of Special Education*, *1*(3), 287-307.

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

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.

Brouwers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher education*, *16*(2), 239-253.

Bruhn, A., McDaniel, S., & Kreigh, C. (2015). Self-monitoring interventions for students with behavior problems: A systematic review of current research. Behavioral Disorders, 40(2), 102-121.

Chafouleas, S. M., Johnson, A. H., Riley-Tillman, T. C., & Iovino, E. A. (2021). School-Based Behavioral Assessment: Informing Prevention and Intervention. Guilford Publications.

Chafouleas, S. M., Riley-Tillman, T. C., Jaffery, R., Miller, F. G., & Harrison, S. E. (2015). Preliminary investigation of the impact of a web-based module on direct behavior rating accuracy. School Mental Health, 7, 92-104.

Chafouleas, S. M., Briesch, A. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2011). Usage Rating Profile–Intervention (Revised). *Storrs, CT: University of Connecticut*.

Chalk, K., & Bizo, L. A. (2004). Specific praise improves on-task behavior and numeracy enjoyment: A study of year four pupils engaged in the numeracy hour. *Educational Psychology in Practice*, *20*(4), 335-351.

Christofferson, M., & Sullivan, A. L. (2015). Preservice Teachers' Classroom Management Training: A Survey of Self-Reported Training Experiences, Content Coverage, And Preparedness. *Psychology in the Schools*, *52*(3), 248-264.

Conroy, M. A., Sutherland, K. S., Snyder, A., Al-Hendawi, M., & Vo, A. (2009). Creating a positive classroom atmosphere: Teachers' use of effective praise and feedback. *Beyond Behavior*, *18*(2), 18-26

Cooper, J. T., Gage, N. A., Alter, P. J., LaPolla, S., MacSuga-Gage, A. S., & Scott, T. M. (2018). Educators' self-reported training, use, and perceived effectiveness of evidence-based classroom management practices. *Preventing School Failure: Alternative Education for Children and Youth*, *62*(1), 13-24.

Corcoran, T. B. (1995). Helping Teachers Teach Well: Transforming Professional Development. CPRE Policy Briefs.

Devine, M., Meyers, R., & Houssemand, C. (2013). How can coaching make a positive impact within educational settings? Procedia-Social and Behavioral Sciences, 93, 1382-1389.

Duchaine, E. L., Jolivette, K., & Fredrick, L. D. (2011). The effect of teacher coaching with performance feedback on behavior-specific praise in inclusion classrooms. *Education and Treatment of Children*, *34*(2), 209-227.

Emmer, E. T., & Stough, L. M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational psychologist*, *36*(2), 103-112.

Faul, A., Stepensky, K., & Simonsen, B. (2012). The effects of prompting appropriate behavior on the off-task behavior of two middle school students. *Journal of positive behavior interventions*, *14*(1), 47-55

Fingerhut, J., Xu, X., & Moeyaert, M. (2021). Selecting the proper Tau-U measure for single-case experimental designs: Development and application of a decision flowchart. *Evidence-Based Communication Assessment and Intervention*, *15*(3), 99-114.

Flood, W. A., Wilder, D. A., Flood, A. L., & Masuda, A. (2002). Peer-mediated reinforcement plus prompting as treatment for off-task behavior in children with

attention deficit hyperactivity disorder. *Journal of Applied Behavior Analysis*, *35*(2), 199-204

Floress, M. T., Beschta, S. L., Meyer, K. L., & Reinke, W. M. (2017). Praise research trends and future directions: Characteristics and teacher training. *Behavioral Disorders*, *43*(1), 227-243

Flower, A., McKenna, J. W., & Haring, C. D. (2017). Behavior and classroom management: Are teacher preparation programs really preparing our teachers?. *Preventing School Failure: Alternative Education for Children and Youth*, *61*(2), 163-169.

Freeman, J., Simonsen, B., Briere, D. E., & MacSuga-Gage, A. S. (2014). Pre-service teacher training in classroom management: A review of state accreditation policy and teacher preparation programs. *Teacher Education and Special Education*, *37*(2), 106-120.

Grasley-Boy, N., Gage, N. A., & MacSuga-Gage, A. S. (2019). Multitiered support for classroom management professional development. *Beyond Behavior*, *28*(1), 5-12.

Grasley-Boy, N. M., Reichow, B., van Dijk, W., & Gage, N. (2021). A systematic review of tier 1 PBIS implementation in alternative education settings. *Behavioral Disorders*, *46*(4), 199-213.

Hamre, & 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. https://doi.org/10.1111/1467-8624.00301

Horne, S. E. (1980). Classroom management. *British Journal of Teacher Education*, *6*(3), 228-235.

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.

Freeman, J., Sugai, G., Simonsen, B., & Everett, S. (2017). MTSS Coaching: Bridging Knowing to Doing. Theory Into Practice, 56(1), 29-37. https://doi.org/10.1080/00405841.2016.1241946

Ingersoll, R. M., & Smith, T. M. (2003). The wrong solution to the teacher shortage. *Educational leadership*, *60*(8), 30-33.

Joyce, B., & Showers, B. (1982). The coaching of teaching. Educational Leadership, 40, 4–10.

Kazdin, A. E. (2011). Single-case research designs: Methods for clinical and applied settings. Oxford University Press.

Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and psychopathology*, *10*(2), 165-185

Kennedy, C., & Jolivette, K. (2008). The effects of positive verbal reinforcement on the time spent outside the classroom for students with emotional and behavioral disorders in a residential setting. *Behavioral Disorders*, *33*(4), 211-221.

Kratochwill, T.R., Hitchcock, J., Horner, R.H., Levin, J.R., Odom, S.L., Rindskopf, D.M. & Shadish, W.R. (2010). What Works Clearinghouse single-case design technical documentation version 1.0. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_scd.pdf

MacSuga-Gage, A. S. (2013). Supporting teachers' professional development: Investigating the impact of a targeted intervention on teachers' presentation of opportunities to respond.

MacSuga-Gage, A. S., & Simonsen, B. (2015). Examining the effects of teacher-directed opportunities to respond on student outcomes: A systematic review of the literature. *Education and Treatment of Children*, 211-239.

Martin, N. K., Schafer, N. J., McClowry, S., Emmer, E. T., Brekelmans, M., Mainhard, T., & Wubbels, T. (2016). Expanding the definition of classroom management: Recurring themes and new conceptualizations. *Journal of Classroom Interaction*, 31-41.

Murray, D. W., Murr, N., & Rabiner, D. L. (2012). Preliminary Effects of the Incredible Years Teacher Training Program on Classroom Management Skills. *Society for Research on Educational Effectiveness*.

Myers, D., Sugai, G., Simonsen, B., & Freeman, J. (2017). Assessing teachers' behavior support skills. *Teacher Education and Special Education*, *40*(2), 128-139.

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.

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), 87-106.

Oliver, R. M., & Reschly, D. J. (2007). Effective Classroom Management: Teacher Preparation and Professional Development. TQ Connection Issue Paper. *National comprehensive center for teacher quality*.

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.

Oliver, R. M., Wehby, J. H., & Nelson, J. R. (2015). Helping teachers maintain classroom management practices using a self-monitoring checklist. *Teaching and Teacher Education*, *51*, 113-120.

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.

Pas, E. T., Bradshaw, C. P., & Cash, A. H. (2014). Coaching classroom-based preventive interventions. Handbook of school mental health: Research, training, practice, and policy, 255-267

Pisacreta, J., Tincani, M., Connell, J. E., & Axelrod, S. (2011). Increasing teachers' use of a 1: 1 praise-to-behavior correction ratio to decrease student disruption in general education classrooms. *Behavioral Interventions*, *26*(4), 243-260.

Poznanski, B., Hart, K. C., & Cramer, E. (2018). Are teachers ready? Preservice teacher knowledge of classroom management and ADHD. *School Mental Health*, *10*, 301-313.

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., 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), 315-332.

Riley-Tillman, T. C., Burns, M. K., & Kilgus, S. P. (2020). *Evaluating educational interventions, second edition : Single-case design for measuring response to intervention*. Guilford Publications.

Royer, D. J., Lane, K. L., Dunlap, K. D., & Ennis, R. P. (2019). A systematic review of teacher-delivered behavior-specific praise on K–12 student performance. *Remedial and Special Education*, *40*(2), 112-128

Scruggs T. E., Mastropieri M. A. (2001). How to summarize single participant research: Ideas and applications. *Exceptionality*, 9(4), 227–244.

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. http://www.jstor.org/stable/42899983

Simonsen, B., Freeman, J., Myers, D., Dooley, K., Maddock, E., Kern, L., & Byun, S. (2020). The effects of targeted professional development on teachers' use of empirically supported classroom management practices. *Journal of Positive Behavior Interventions*, *22*(1), 3-14.

Simonsen, B., MacSuga-Gage, A. S., Briere III, D. E., Freeman, J., Myers, D., Scott, T. M., & Sugai, G. (2014). Multitiered support framework for teachers' classroom-management practices: Overview and case study of building the triangle for teachers. *Journal of Positive Behavior Interventions*, *16*(3), 179-190.

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

Sims, W. A. (2016). *Evaluating the Effects of Performance Feedback Using the Direct Behavior Rating-Classroom Management (DBR-CM) on Teacher Classroom Management Behavior*. University of Missouri-Columbia

Sims, W. A., King, K. R., Reinke, W. M., Herman, K., & Riley-Tillman, T. C. (2021). Development and preliminary validity evidence for the direct behavior rating-classroom management (DBR-CM). *Journal of Educational and Psychological Consultation*, *31*(2), 215-245.

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

Sutherland, K. S., Wehby, J. H., & Copeland, S. R. (2000). Effect of varying rates of behavior-specific praise on the on-task behavior of students with EBD. *Journal of Emotional and Behavioral Disorders*, *8*(1), 2-8.

Thompson, M. T., Marchant, M., Anderson, D., Prater, M. A., & Gibb, G. (2012). Effects of tiered training on general educators' use of specific praise. *Education and Treatment of Children*, *35*(4), 521-546.

Walker, H. M., Nishioka, V. M., Zeller, R., Severson, H. H., & Feil, E. G. (2000). Causal factors and potential solutions for the persistent underidentification of students having emotional or behavioral disorders in the context of schooling. *Assessment for Effective Intervention*, *26*(1), 29-39.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of educational research*, *63*(3), 249-294.

Wilder, D. A., & Atwell, J. (2006). Evaluation of a guided compliance procedure to reduce noncompliance among preschool children. *Behavioral Interventions: Theory & Practice in Residential & Community-Based Clinical Programs*, *21*(4), 265-272.

Appendix B: DBR-CM SR

Direct Behavior Rating – Classroom Management: Self-Report Form (DBR-CM SR)

Date:	Educator Name:	Observation Start Time:	Instructional topic:
M T W TH F	Observer Name:	End Time:	

Classroom Structure – Classroom, desks, furniture, materials, and technology are arranged in a manner that allows for movement within the classroom without disruption and for the students and educator to easily see one another.

Yes	Somewhat	No
☐	☐	☐

Place a mark in the box that corresponds to your rating for each behavior domain.

Praise
Using positive statements or actions in response to student performance.

0 1 2 3 4 5 6 7 8 9 10
Low Medium High
Rate the level of praise used during the observation period.

Communication
Clearly presenting goals and expectations.

0 1 2 3 4 5 6 7 8 9 10
Low Medium High
Rate the amount of clear communication present during the observation period.

Enthusiasm
Instruction is presented in an accurate, meaningful, memorable, and/or engaging manner.

0 1 2 3 4 5 6 7 8 9 10
Low Medium High
Rate the amount of enthusiasm used to deliver content in an engaging manner during the observation period.

Rapport
The student-educator relationship is mutually positive and accepting.

0 1 2 3 4 5 6 7 8 9 10
Low Medium High
Rate the amount of rapport, or trust, acceptance, and positivity, during the observation period.

Appendix C: DBR-CM ER

Direct Behavior Rating – Classroom Management: External Rater Form (DBR-CM ER)

Date:	Educator Name:	Observation Start Time:	Instructional topic:
M T W TH F	Observer Name:	End Time:	

Classroom Structure – Classroom, desks, furniture, materials, and technology are arranged in a manner that allows for movement within the classroom without disruption and for the students and educator to easily see one another.

Yes	Somewhat	No
☐	☐	☐

Place a mark in the box that corresponds to your rating for each behavior domain.

Praise
Using positive statements or actions in response to student behavior.

0	1	2	3	4	5	6	7	8	9	10
Low					Medium					High

Rate the amount of praise used during the observation period.

Communication
Clearly presenting goals and expectations for an instructional period.

0	1	2	3	4	5	6	7	8	9	10
Low					Medium					High

Rate the amount of clear communication present during the observation period.

Enthusiasm
Instruction is presented in an accurate, meaningful, memorable, and/or engaging manner.

0	1	2	3	4	5	6	7	8	9	10
Low					Medium					High

Rate the amount of enthusiasm used to deliver content in an engaging manner during the observation period.

Rapport
The student-educator relationship is mutually positive and accepting.

0	1	2	3	4	5	6	7	8	9	10
Low					Medium					High

Rate the amount of rapport, or trust, acceptance, and positivity, during the observation period.

Appendix D (DBR-CM ER Example PF Graph: Phase B and Phase C)