DEPENDABILITY AND RELIABILITY OF THE DIRECT BEHAVIOR RATING - CLASSROOM MANAGEMENT FORMS

A Dissertation

Submitted to the Graduate Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the requirements for the degree of Doctor of Philosophy in The Department of Psychology

by

Kaitlin Angela Cassidy
B.A., Loyola University Maryland, 2015
M.A., Louisiana State University, 2018
August 2021
# TABLE OF CONTENTS

Abstract .................................................................................................................................................. iii
Introduction ........................................................................................................................................... 1
Methods .................................................................................................................................................. 28
Results ................................................................................................................................................... 36
Discussion ............................................................................................................................................... 41
Appendix A. Direct Behavior Rating-Classroom Management External Rater Form......................... 47
Appendix B. Direct Behavior Rating-Classroom Management Self-Report Form............................ 49
Appendix C. Rater Training Protocol ................................................................................................... 51
Appendix D. Institutional Review Board Approval Documentation .................................................. 56
References ............................................................................................................................................... 57
Vita .......................................................................................................................................................... 66
ABSTRACT

Classroom management skills are evidence-based strategies used to maintain a positive and productive learning environment. Utilizing classroom management strategies has been proven effective; however, limited tools are available for assessing such skills and further assessment of the reliability and dependability of such measures is needed. The Direct Behavior Rating – Classroom Management (DBR-CM) forms are measures for assessing classroom management and come in three forms: external rater (DBR-CM ER), self-report (DBR-CM SR), and student rater (DBR-CM ST). This study extended on the information obtained from the Cassidy (2018) study that examined the DBR-CM ER form. The study evaluated the inter-observer agreement among the DBR-CM ER and DBR-CM SR forms. Additionally, this study examined the variance contributing to the classroom management ratings for both the DBR-CM SR and DBR-CM ER forms. A fully crossed analytic design (p x d x m) with two facets, day (d: observations) and method (m: rater type), and person (p: teachers) as the object of measurement was created. A follow-up decision (D) study was conducted to assess the number of 10 minute observations required to conduct a reliable rating with the DBR-CM SR and DBR-CM ER forms.
INTRODUCTION

Classroom management skills are class-wide strategies used to maintain a productive learning environment. Further, classroom management can be defined as “all those actions (and conscious inactions) teachers…engage in to enhance the probability that children, individually and in groups, will develop effective behaviors that are personally fulfilling, productive, and socially acceptable,” (Walker, Shea, & Bauer, 2004, pp. 7). General research consensus indicates behavior management must be attained first to succeed in fostering a productive classroom environment (Emmer & Sabornie, 2015) and evidence-based classroom management skills can be utilized to maintain such an environment to ensure optimal access to academics.

Early Twentieth Century Classroom Management

At the onset of the twentieth century William Bagley (1907) sparked an interest in purposeful strategies for managing classrooms beyond corporal punishment by developing a list of managerial principles for the classroom. Bagley derived his managerial principles from a combination of unsystematic teacher observations, personal teaching experience, non-evidence-based education textbooks, and general psychological principles tested within different contexts. For example, Bagley advised for explicit instructions, as well as consistent routines and expectations, to be explained in detail to students on their first day of school. Bagley also discussed fading these detailed instructions and expectations as students exhibit more individual responsibility. While, Bagley was not opposed to corporal punishment, he promoted positive, reward-
based techniques that fostered motivation and encouragement. Although Bagley did not support his claims with research, many of his advice was later researched and can still be found in the literature today.

After Bagley’s essay on classroom management, little attention was cast on the subject and systematic research pertaining to classroom management did not primarily begin until the 1950s. Professionals during the 1950s and 1960s blended influences from educational psychology, the mental health movement, and social psychology to assess classroom climate and teacher leadership styles. Much of this research stemmed from Lewin, Lippitt, and White’s (1939) study concerning leadership and social climate among boy scouts. Lewin et al. (1939) examined the effectiveness of democratic, authoritarian, and Laissez-faire leadership styles among 10 year-old boy scout groups. This research indicated a democratic leadership style is the most responsive and independently productive leadership style. The leader within the democratic style was objective and fact-minded in the criticism and praise of members. The democratic leadership style also involved group discussion and decision regarding all policies matters, with assistance from the leader. Additionally, the members were able to work with whomever they chose and divide tasks as they chose.

In order to expand these findings and examine effective leadership styles of teachers, researchers began to employ the use of systematic classroom observations (Simon & Boyer, 1970). For instance, Ryans (1952) sent several observers to assess elementary and secondary classrooms. These observers were trained to utilize the Classroom Observation Scale, which observes teacher and pupil behaviors. This study
concluded teacher management is positively correlated with sociability, fairness, responsibility, consistency, student engagement, enthusiastic, open-minded, as well as the superficial appearance of the teacher.

By the 1970s it was widely accepted that positive, reward-based strategies, paired with an authoritarian leadership style was the preferable managerial method within the classroom; therefore, researchers interested in classroom management began examining more specific components of effective classroom management skills. This movement was taken up by two separate sources: behaviorists and ecological researchers. Behavioral research regarding classroom management has largely featured applied behavior analysis principles, especially operant conditioning and reinforcement schedules implemented to shape behavior (Brophy, 2006). Early implementation of behavioral techniques derived from laboratory-generated findings and were successfully modified to fit within the classroom context (Brown, 1971; Brophy, 2006).

Ecological research examined environmental contexts with an emphasis on supported and prohibited activities (Bronfenbrenner, 1989) governed by teachers, peers, other adults’ influence, as well as the physical setting (Brophy, 2006). A prominent ecological researcher, Kounin conducted several studies with colleagues examining effective classroom management. For instance, Kounin and Gump (1958) observed the bystander reactions to another student’s discipline within the classroom. Teachers’ responses to disruption minimized further disruption by audience peers when the response was firm and explicit, however, it failed to impact similarly disruptive peers when the teachers responded harshly. Kounin’s other studies (Kounin, Friesen, &
Norton, 1966; Kounin, 1970; Kounin & Doyle, 1975) indicated a teacher’s overall effectiveness in decreasing disruptive behaviors lies in proactive strategies to maintain activity momentum and nipping potential problems at their start. Researchers, such as Emmer and Evertson, built from Kounin’s research and concluded teachers who are effective classroom managers demonstrate mastery within three clusters of behaviors: conveying purposefulness, teaching appropriate conduct, and sustaining student attention (Brophy, 2006).

**Current Classroom Management Literature**

Current literature supports the benefits of evidence-based classroom management strategies and suggests utilizing a combination of both reactive and preventative classroom management procedures (Doyle, 1986; Evertson & Emmer, 1982; Jones & Jones, 1986). Extensive classroom management includes proximate responding to students’ current behaviors and preemptive planning for explicitly stating expectations and praising productive behaviors. Proactive management purposefully minimizes opportunities for disruptive behaviors. Consequentially, proactive management decreases the time devoted to student discipline (Gettinger, 1988) and therefore, increasing the time allotted to class-wide academic instruction. Common strategies involve active student engagement, consistent active teacher supervision, clear student performance feedback and frequent, specific praise of student behavior (Gettinger, 1988; Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008).

Optimally managed classrooms are characterized by the teacher’s ability to monitor student compliance and attention, establish explicit behavioral expectations, and implement procedures to proactively prevent a disruptive atmosphere (Gettinger,
Accordingly, well-managed classrooms exhibit low levels of conflict and disruptive behavior, high levels of respectful communication and problem solving, strong on-task focus, smooth transitions, appropriate emotional expressions, and sensitivity to student differences and needs (Pianta & La Paro, 2003). Consistent implementation of these strategies can also result in increased academic engagement, reduced classroom interruptions, and minimal individual student disruptions (Brophy, 1985; Kounin, 1970; Simonsen et al., 2008).

Conversely, a teacher’s difficulties in implementing classroom management skills are correlated with lower levels of academic performance and on-task behavior among students (Reinke et al., 2008). This can also lead to the absence of positive teacher-student relationships, unclear expectations, and student academic failure and behavioral problems (Gottlieb & Plirstock, 2005; Split, Hughes, Wu, & Kwok, 2012). To correct this potential trend, classroom management strategies should be examined for the purpose of identifying teachers who exhibit difficulties with managing their classrooms and areas for improvement among those teachers. The presentation of assessment data (e.g., performance feedback) has a significant positive impact on performance improvement (Fuchs, Fuchs, Hamlett, & Stecker, 1991); therefore, providing easy and affordable access of this type of data to teachers and administrators should help inform areas for improvement and identify at-risk classrooms.

Classroom Management Support

Although evidence-based classroom management strategies (e.g., clearly formulated rules, consistent routines, and efficient organization) are commonly identified as effective tools in increasing classroom engagement and academic success (Kunter,
Baumert, & Köller, 2007; Praetius, Pauli, Reusser, Rakoczy, & Klieme, 2014), teachers often obtain little classroom management support or formal training (Simonsen et al., 2020, Stough & Montague, 2015) especially in regards to behavior management of students with severe behavior problems (Wagner et al, 2006). For example, Martinussen, Tannock, & Chaban (2011) found 75% of general education teachers and 46% of special education teachers surveyed reported having no or limited in-service training related to ADHD and evidence-based behavior management of children with ADHD. Additionally, classroom management as an independent course is inconsistently accessible in teacher training/university programs (Brophy, 2006; Stough & Montague, 2015) and as such, teachers often enter their classrooms with minimal pre-service training related to behavior management in the classroom (Freeman, Simonsen, Briere, & MacSuga-Gage, 2014).

Further, efforts to integrate classroom management strategies within professional development workshops are frequently more instructional than applied (Guskey, 2000). This method is often ineffective due to shortcomings in being able to acquire and master skills within a one-session training (Sugai et al., 2000). Further, even though classroom management strategies have been well established, there is an existing gap between knowledge and implementation of behavior management practices in the classroom (Simonsen et al., 2020).

As previously mentioned, student academic success is associated with better managed classrooms. A deficiency in consistent and effective managerial skills may lead to student academic failures and behavioral problems, consequentially producing a surplus in unnecessary student referrals through the Response to Intervention (RTI)
process and perchance inappropriate placement in special education (Gottlieb & Plirstock, 2005). To ensure confidence in the RTI process or systems devoted to special education placement, evidence-based classroom management techniques should be applied to diminish the prospect of inadequate classroom management tactics which may incidentally reinforce negative social behaviors or academic performance in students. Therefore, by having an effective tool to recognize teachers in need of management strategies, there could be a decrease in the number of students unnecessarily referred to administration for concerns and an increase in effective Tier I intervention plans. Hence, evidence-based tools are necessary to identify which classrooms need effective implementation of classroom management strategies, in addition to progress monitoring the implementation of these skills.

Teachers identified with managerial deficiencies at the onset of a school year often have continuing disciplining difficulties throughout the year (Emmer & Evetson, 1981) and early identification and implementation of classroom strategies could combat these difficulties. McCarthy, Lineback, & Reiser (2015) suggest that targeting teachers’ deficits in classroom management could help increase self-efficacy, as well as help them manage their classrooms more effectively. Regular performance assessment could allow for identification of teachers in need of more classroom training support. Currently, there are minimal psychometrically sound tools for assessing classroom management easily, efficiently, and affordably for the purposes of skill-level identification and constructive feedback (Simonson et al., 2020). If such tools were widely known and available, a problem-solving multitier framework could be employed to efficiently identify and appropriately train at-risk teachers.
Adequate preparation for classroom management may benefit teachers to the same degree that it benefits student wellbeing and academic success. For instance, teachers with classroom management training are more likely to utilize effective strategies (Hall & Wahrman, 1988), effectively respond to aggressive students (Alvarez, 2007), and feel more confident in their abilities to handle misbehavior (O’Neill & Stephenson, 2012). Additionally, previous research suggests classroom management training may decrease teacher stress, burnout, and attrition rates (Ingersoil & Smith, 2003; O’Neill & Stephenson, 2012).

Classroom Observation Tools

There has recently been a push for professional development trainings to be developed within a multitiered system of supports framework (Simonson et al., 2020; Freeman et al., 2017); however, in order to appropriately customize behavior management trainings within a tiered system, school leaders require reliable data to guide decision making and therefore, also require reliable tools to collect this data.

There are several classroom management checklists (e.g., Classroom-Management Assessment [CMA], 2011) and brief classroom walk-through tools (e.g., Missouri School-Wide Positive Behavior Support [MO SWPBIS], 2017); however, there is limited research related to their psychometric properties (Simonsen, 2020). Checklists can be completed within the classroom or outside of the classroom and consist of a series of yes or no response items; these types of measures are limited in feedback to teachers. A walk-through tool is a specific type of observation assessment that consists of a checklist or Liker-scale measure that is expected to examine just a snap shot of class time.
Classroom observations are often the chosen method for examining classroom instruction quality, as well as student or teacher behavior (Praetius, et al., 2014). Common essential features of behavior assessment often considered when selecting a tool include feasibility, defensibility, flexibility, efficiency, and repeatability (Chafouleas, Volpe, Gresham, & Cook, 2010). Several of the current approaches to assess teachers' classroom management are either psychometrically sound and resource intensive, or efficient but psychometric properties are unknown (Simonsen, 2020). For example, The Classroom Assessment and Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008) is a frequently cited and utilized observation tool with extensive established psychometric properties; however, it is very resource-intensive. In order to utilize the CLASS, observers must complete a certification course, routinely renew their certification, as well as routinely purchase the observation assessment forms and manuals. Further in order to establish a reliable assessment of behavior management in a classroom, the observer is required to conduct six 30 minute observations.

In contrast, several observation measures, such as the Classroom Management Observation Tool (CMOT; Simonsen, 2019), are simple, efficient, and freely available online; however, their research regarding their psychometric properties is often limited. The CMOT requires a single 15 minute observation; however, it is a very recently introduced measure and requires further research to develop the tool for screening, progress monitoring, and other assessment purposes (Simonsen, 2020). Further a generalizability study examining the reliability of the CMOT supported score differences for items related to active supervision and response opportunity were largely due to
teacher differences; however, the majority of variance remained unexplained for items related to specific praise and the favorable positive to corrective ratio (Simonsen, 2020).

**Behavior Rating Scales**

Behavior rating scales are commonly utilized within the school setting today. For instance, Shapiro & Heick (2004) reported 75% of school psychologists reported utilizing behavior rating scales when evaluating emotional-behavioral referrals. Behavior rating scales are measures completed by an informed respondent (e.g., administrator, supervisor, self, etc.) for the purpose of rating or estimating the target individual across varying behavior domains (Martin, Hooper, & Snow, 1986). These ratings are typically based on past anecdotal or informal observations of that individual over a specified timeframe (Campbell & Hammond, 2014). Behavior rating scales are routinely utilized for assessment data collection because they are psychometrically sound tools that aid in identifying one’s strengths and weaknesses. Behavior rating scales also allow for collecting data from various sources and settings, developing effective services, and monitoring the individuals progress and outcomes (Campbell & Hammond, 2014). Further, rating scales are often low-cost, capable of assessing low-frequency or private behaviors, capable of addressing a wide breath of behavior problems, often include normative data samples to interpret findings, and are more objective than interviewing techniques (Whitcomb & Merrell, 2013; Volpe, McConaughy, & Hintze, 2009).

Behavior rating scales within the school system are most used to assess individual students’ behaviors; however, they have been utilized to assess teachers’ performance and other related behaviors and are often completed by an administrator or the teachers themselves. For example, behavior rating scales have been utilized to
assess teachers’ stresses, and behavior management strategies. Most commonly these forms are used when rating teacher job performance. While there are various advantages to utilizing behavior rating scales in this manner, a major disadvantage is that behavior rating scales are estimates of behavior rather than a direct account of behavior (Campbell & Hammond, 2014). Estimates of behavior can be concerning as they may lead to biased under or over-reporting of behaviors. Therefore, while these scales can be helpful in making informed opinions, it is advantageous to include direct observations to assess behaviors. Other limitations of behavior rating scales include being occasionally costly and increased risk of over-generalizing or misinterpreting data collected from the rating scales (Whitcomb & Merril, 2008).

**Systematic Direct Observations**

Systematic direct observation (SDO) is often regarded as the standard for reliable and accurate measurement of target behaviors within a classroom (Cone, 1978; Riley-Tillman, Christ, Chafouleas, Boice-Mallach, & Briesch, 2011). An SDO measure is an observation procedure that allows for quantifying target behaviors in a standardized manner by codifying occurrences and non-occurrences of the target behaviors (Hintze, Volpe, & Shapiro, 2002; Ferguson, Briesch, Volpe, & Daniels, 2012). Specific time intervals are often pre-determined to allow for a standard process of scoring and summarizing observation findings; additionally, they can be easily adapted across contexts and observers. School psychologists and other school professionals often utilize SDOs to inform decision-making processes within school settings (Ferguson, et al., 2012).
Advantages of utilizing an SDO include being a form of direct assessment where real-time behaviors are measured and being multi-purposeful (e.g., screening, diagnosing, treating, and monitoring). Additionally, it allows for a narrow scope of targeted behaviors by requiring operational definitions covering the topography, magnitude, and function of the behaviors being observed. While SDOs have strengths in inter-observer agreement and sensitivity to behavior change (Briesch, Chafouleas, & Riley-Tillman, 2010), this technique has been criticized for infeasibility. SDOs are time sensitive and laborious, while also often requiring excessive rater training (Riley-Tillman, Kalberer, & Chafouleas, 2005; Riley-Tillman, et al., 2011). Additionally, SDOs focus on detailed descriptions of target behaviors requiring a limited breadth of behaviors that can be concurrently observed. Further, threats to the validity of SDOs include poorly defined behaviors, observer drift, recording errors, biased observers, and no available normative comparisons (Volpe, McConaughy, Hintze, 2009; Hintze, Volpe, & Shapiro, 2002).

**Direct Behavior Ratings**

A more feasible technique for assessing classrooms are direct behavior rating (DBR) scales. DBR scales incorporate the ease of traditional rating scales as well as the positive features of SDO measures (Riley-Tillman, et al., 2011). DBR scales are created by organizing a group of categorically related behaviors into single measurable domains, which are then assessed on frequency or duration using a Likert-like scale (Deno, 2003). DBR ratings are often completed by individuals with high familiarity of the measured target behaviors and the DBR ratings are recorded in close temporal proximity to the assessed target behaviors (Christ, Riley-Tillman, Chafouleas, & Boise,
DBR scales inform explicit target behavior definitions, provide examples and non-examples of target behaviors, allow for prompt ratings, and require minimal classroom interference (Christ, Riley-Tillman, & Chafoules, 2010). Additionally, the instrumentation and procedures of a DBR tolerate vast flexibility in potential target behaviors; thus, it is an easily accessible solution to assessing many school related concerns, such as assessing classroom management (Chafouleas, et al., 2013; Sims, 2016). The combination of these features builds on the strengths of both SDO and behavior rating scales (Christ et al., 2010), while also accounting for the weaknesses of these types of observation measures (Riley-Tillman et al, 2005).

Typically, DBR scales rate the frequency or duration of specific target behaviors regarding individual students. While this is valuable in assessing specific children at risk for behavior concerns, a classroom measure examining teacher management skills, would allow school psychologists or administrators the ability to identify classrooms with poor classroom management and therefore at risk for problematic behaviors. This approach would allow for focusing on Tier I intervention implementation, rather than emphasizing Tier III interventions. When DBR scales are utilized for individual students, typically it would lead to individualized interventions targeting that student behaviors; however, utilizing a DBR scale that focuses on the classroom environment would lead to improvements in universal behavior management strategies which may help limit the need for individualized plans. When teachers can implement universal strategies instead of multiple individualized plans, they have more time to devote to instruction. Currently, school administrators observe teachers infrequently (Weisberg et al., 2009).
and often via non-descript rubrics that focus on global standards without providing specific feedback for improving their skills (Danielson, 2013).

**Teacher Self-Report**

Regular and specific performance assessment by an external rater allows for identification of teachers in need of more classroom training support; however, if a self-report measure can also yield reliable results of classroom management strategies, it could mean a more efficient progress monitoring tool for teacher progress. Depending on external raters (e.g., administration staff) to conduct initial screeners as well as progress monitoring assessments of classroom management skill-level is time intensive and costly, as well as potentially unnecessary. Training teachers to effectively self-monitor their own behaviors and progress would be much more sufficient in regard to time, money, and resources spent. Previous research has identified the two most effective adult learning strategies as self-evaluation and self-identifying performance goals (Dunst, Trivette, & Hamby, 2010). Further, positive effects to self-monitoring may be attributed to the fact that a self-monitoring sheet can install reactivity and prompting to implement learned evidence-based strategies (Oliver, Wehby, & Nelson, 2015), rather than relying on previous non-effective habits. It is likely that training teachers on the assessment can indirectly serve as self-provided constructive feedback and a proactive prompt for effective classroom management tools; ultimately creating a dual purpose for a self-report measure, as an assessment tool (i.e., self-report measure) and as an intervention permanent product (i.e., self-monitoring intervention method).

The first step to determining the feasibility of utilizing a self-report DBR for the purposes described above, would be to determine the accuracy and dependability of
self-report scores. Unfortunately, self-reports are known to produce severe bias, often due to situational factors, individual’s mood, lapses in memory, or social desirability effects (Diener, 2000). Social desirability bias refers to one’s tendency to endorse more socially desirable behaviors within certain contexts (Richman, Weisband, Kiesler, & Drasgow, 1999; Caputo, 2017). An advantage of self-report observations rather than many self-report behavior rating scales, is that these identified limitations can be in part examined by evaluating the discrepancy between concurrent teachers’ self-report and a trained professional who completes an external raters’ observation form. Additionally, teacher self-reports have demonstrated reliability and validity when forms are behavior specific, time specific, and completed multiple times (Newfield, 1980; Koziol & Burns, 1986). Although external rating forms may have higher reliability than self-reports, it is hypothesized that with efficient training a self-report form assessing classroom management skills could maintain adequate reliability and be utilized to self-monitor teachers’ skill level in applying management strategies.

**Direct Behavior Rating-Classroom Management**

The Direct Behavior Rating – Classroom Management (DBR-CM; Sims, 2014) forms measure classroom management skills and can be utilized for the purposes discussed above. It is a classroom management assessment developed to be a feasible, flexible, and defensible classroom management tool (Sims et al., 2021). The DBR-CM intends to provide screening and progress monitoring data to identify and support educators behavior management development (Sims et al., 2021). This measure combines the strengths of rating scales and systematic direct observations to produce a measure that is feasible, efficient, and defensible. This measure is easily
accessible online at no cost, requires brief training, and can be completed within a time frame of the observers choosing.

There are three versions of measure: external rater (ER), student rater (ST), and self-report (SR); however, this study focuses on the DBR-CM ER and DBR-CM SR forms. Each of these measures contain the same exact items and protocol information, the only difference being the person expected to complete the form. The DBR-CM forms add to the limited number of tools available for utilizing a problem-solving oriented approach to rate teacher classroom management performance (Sims et al., 2021). The DBR-CM forms are composed of five subscales: classroom structure, praise, communication, enthusiasm, and rapport. Operational definitions and specific behavior examples and non-examples, for all the domains, are supplied on the back of the DBR-CM forms and are based on the current classroom management research literature. Similar to the format of typical DBR forms, the DBR-CM forms utilize a Likert-scale rating system, ranging from 0 to 10.

Classroom Structure. Classroom structure refers to whether the furniture and other materials or supplies in the classroom are organized in a manner that allows students and teachers to easily see each other and move about the room, without adding additional disruption (Sims, 2014). Unlike all of the other domains, Classroom Structure is not measures utilizing a Likert-scale rating; instead, the rater is expected to respond Yes, Somewhat, or No to whether the classroom meets criteria for a well-structured learning environment.
Praise. Teachers’ specific praise increases student motivation, helpful and positive feedback, and positive student-teacher relationships (Bear, 2015). The DBR-CM forms, operationally define praise 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. 36). Positive praise statements are ideally behavior-specific, though general praise applies as well (Sims, 2016). General praise can be a reinforcer that is verbal, gestural, or tangible (Sims, 2014). For instance, if a teacher awards points or a thumbs-up for appropriate behavior without a verbal explanation, it would qualify as general praise. If the gesture or tangible reward were accompanied with a verbal explanation for the reinforcement, it would qualify as behavior specific praise. Teachers who receive a high rating for praise should reward students with specific praise at a higher frequency than general praise (Sims, 2016). As per the operational definition for this study, the student must be mindful of receiving a tangible reward for it to be considered praise and praise statements should be contingent on expected behaviors (Sims, 2016). Further, it is widely accepted to provide three or four praise statements for every reprimand given (Epstein, Atkins, Cullinan, Kutash, & Weaver, 2008; Sims, 2016). Therefore, this domain expects the number of reprimands to not exceed the four to one ratio. When reprimand is necessary, it should be quick and, in a calm, non-harsh tone (Sims, 2014). This construct also includes displaying a more positive than negative attitude and tone during student interactions (Sims, 2016).

Communication. Clear and explicit communication of rules and expectations is an essential component of classroom management, as it decreases potential student
misconduct (Ratcliff, 2001). The following construct, communication, is operationally defined as “clearly conveying goals and expectations of a classroom and/or instructional period to students,” (Sims, 2016, p. 36). Communication includes verbally and/or visually delivering explicit behavioral expectations and academic objectives, as well as ensuring time for addressing questions (Sims, 2016). It is important that these behavioral expectations and academic objectives are age appropriate and easily communicated for all students. For instance, if a teacher has a list of classroom rules, they should be posted, with age appropriate vocabulary, where the whole class can clearly see it. Additionally, teachers should communicate the behavior they expect of their students rather than asking students to stop inappropriate behavior. For example, teachers should tell students, “to place their bottoms on their chair and feet on the floor” rather than telling students, “not to sit on their knees.” Further, good communication is demonstrated by many students knowing and obeying transition routines and attention signals (Sims, 2016; Sims, 2014).

_Enthusiasm._ Positive student views of teacher enthusiasm and attentiveness are positively related to students’ academic enjoyment and intrinsic value of the academic material (Keller, Goetz, Becker, Morger, & Hensley, 2014). The next construct, enthusiasm, is operationally defined as “the delivery of instructional content in a meaningful, memorable, and/or engaging manner” (Sims, 2016, p. 38). Four or more questions should be posed per minute, with most students answering at least one question during instruction (Sims, 2016). Moreover, minimal behavioral disruptions should be observed throughout the instruction (Sims, 2016). Teachers should provide
accurate instruction in a positive and upbeat tone, while utilizing an appropriate pace (Sims, 2014). Additionally, real-world examples, multiple modalities, and alternative activities (e.g., group work, current events, students teaching students, etc.) should be incorporated to supplement learned material (Sims, 2014).

Rapport. Student-teacher relationships significantly impact students social and academic development (Hughes, Luo, Kwok, & Lyod, 2008; Hamre & Pianta, 2001; Mercer & DeRosier, 2008). The last construct, rapport, is operationally defined as “the quality of the student-teacher relationship, especially that of mutual trust, emotional affinity, acceptance and positivity” (Sims, 2016, p. 37). Rapport consists of reciprocated feelings of warmth and acceptance between the teacher and student (Sims, 2016). Teachers demonstrate good rapport by frequently referring to children by their names, as well as appearing sincere, encouraging, and calm in their exchanges with students (Sims, 2016). During these interactions, the teacher and students are both visibly comfortable and upbeat; students visibly seem comfortable approaching the teacher with questions, comments, or other statements (Sims, 2014).

Reliability and Dependability of DBR-CM Forms

Initial reliability and validity data (Sims et al., 2021) suggest that the DBR-CM is psychometrically sound. For instance, initial validation efforts provide support for concurrent validity (Sims et al., 2021). The DBR-CM variables are significantly positively associated with the Brief Classroom Interaction Observation-Revised (BCIO-R). Specifically, the DBR-CM domains (i.e., enthusiasm, praise, communication & rapport) were found to be significantly positively correlated with the BCIO-R variables (i.e., rate of OTR, rate of overall praise, and rate of overall reprimands). Additionally, the DBR-CM
Praise and Rapport items are positively correlated with the BCIO-R Positive Implementation variable. Further, the DBR-CM was also found to significantly positively correlated with the Classroom Atmosphere Scale (CAS) and the Ohio State Teacher Efficacy Scale (OSTES). The DBR-CM appears to be significantly positively correlated with the CAS total score and OSTES total scores. Further, all individual DBR-CM and CAS items were found to be significantly positively correlated to varying strengths (Sims et al., 2021).

Sims and colleagues (2021) found inter-rater reliability statistics approached or exceeded 70%. Specifically, DBR-CM Praise (69%), Engagement (69%), and Rapport (67%) were found to approach 70%; whereas Communication (75%), Enthusiasm (78%), and Total (70%) met or exceeded 70%. Utilizing a similar training method, Cassidy (2018) found inter-observer agreement for DBR-CM ER Total (87%), Communication (93%), Rapport (86%), and Enthusiasm (84%) exceeded 80%, whereas and Praise equaled 70%. Further research, is required to further assess this measures IOA, given the discrepancy in results from these two studies.

Cassidy (2018) also examined the generalizability of ratings produced, where results suggested the majority of the variance (49%) among scores can be attributed to the actual teacher being observed. The G study also produced a significant G coefficient (.91) and a significant Phi coefficient (.91), indicating the DBR-CM ER assessment produces sufficient relative & absolute dependability of ratings measuring classroom management skills. The G and Phi coefficients were utilized within a follow-up D-Study to determine the length of time needed, as well as the number of observations required to conduct a reliable rating with the DBR-CM ER. The results
suggested for the most efficient use of the measure, a trained rater can observe a classroom four times, for a ten minute interval each observation, to sufficiently produce reliable estimates of a teachers classroom management skills.

**DBR Limitations**

The greatest limitations of DBR style scales are rater error (Riley-Tillman, Chafouleas, Christ, Briesch, & LeBel, 2009; Briesch, Chafouleas, & Riley-Tillman, 2010) and variance in observation durations (e.g., length and number of observations; Ferguson et al., 2012).

*Rater Error and Training.* Although the feasibility, repeatability, and familiarity of DBRs indicate high potential (Riley-Tillman et al., 2009), there is concern about the influence of rater error. Riley-Tillman, et al. (2009) utilizes the term rater error to label situations where ratings tend to either over- or underestimate the observed target behavior’s true score. Additionally, Briesch et al.’s (2010) results imply rater-related effects largely contribute to the error variance estimates of DBR ratings. Since behavioral assessments should contain trustworthy scores, this concern needs addressing.

Error variance in DBR ratings may be affected by the absence of systematic rater training among previous research studies (Chafouleas et al., 2013). Fortunately, brief DBR rater trainings involving practice and feedback have improved rater accuracy in studies that have included a training component (Harrison, Riley-Tillman, Chafouleas, 2014; Schlientz, Riley-Tillman, Briesch, Walcott, & Chafouleas, 2009). Therefore,
confidence in rater accuracy of the DBR-CM forms may be improved by conducting a routine rater training.

Sims and colleagues (2016; 2020) implemented a training procedure for the DBR-CM ER where procedures involved a presentation of operational definitions, rating methodology, and video examples and non-examples of classroom management. Research assistants then practiced the learned material by watching and rating example videos of teacher classroom management. Cassidy (2018) employed a similar training procedure and within their study; however, as discussed above, there were discrepancies among IOA percentages between the two studies. Establishing inter-observer agreement (IOA) in measures of behavior is essential for guaranteeing a level of dependability and consistency among observers (Westling, Koorland, & Tait, 1981).

Duration of Observation Assessments. Variance in observation durations (e.g., length and number of observations; Ferguson et al., 2012) is also a major concern for the utilization of DBR forms. Fortunately, previous research has examined the necessary observation duration needed for dependable ratings of the DBR-CM ER (Cassidy, 2018). In Cassidy (2018), the generalizability theory analyses determined the DBR-CM ER produces reliable ratings within four, ten minute intervals. No such research has yet to be conducted on the DBR-CM SR. Based on the Cassidy (2018) findings, this study presumes a ten minute observation timeframe will also be sufficient for the DBR-CM SR. Given the lack of feasibility in asking teachers to pay close attention to and rate their classroom management strategies across several timeframes (e.g., after each 5- minute interval), this study will only be examining the number of
observations required to determine dependable scores rather than also examining the required length of those observations as well.

**Generalizability Theory**

The classical test theory (CTT) is the common approach for assessment analyses. CTT is valuable in comprehending the degree of accuracy with which measurements are conducted, as well as the overall strength or weakness of a measure (Briesch, et al., 2010); however, CTT does not produce information about the means for strengthening the measurement or reducing error within the measurement (Briesch, et al., 2010). An alternate approach to analyzing the psychometric properties of a direct behavior rating scale is Generalizability Theory (G Theory) analyses. Unlike CTT, G theory does generate information regarding how to improve a measurement, rather than simply indicating the overall strength and weakness of a measure (Briesch et al., 2010; Cone, 1978; Gresham & Carey, 1988).

G Theory is a statistical framework, which tests the dependability of a measure by computing data about the reliability and validity of a measure (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Suen, 1990; Hintze & Matthews, 2004). Unlike CTT, G theory allows for multiple sources of error variance to be partitioned, and therefore concurrently analyzed via a G Study. This partitioning allows researchers to identify specific sources of measurement error (i.e., facets and the interactions between facets) that are of concern, and then evaluate the relative extent of each of those sources (Hintze & Matthews, 2004). In addition to variance scores, Phi and generalizability (G) coefficients will be calculated to examine relative and absolute dependability of the data,
respectively. Phi coefficients are similar to Pearson correlation coefficients and estimate the extent of a relationship between two variables. G coefficients are similar to reliability coefficients in CCT and estimate the ratio of universe-score variance to the expected observed-score variance. Ideally, a G study will produce phi and generalizability (G) coefficients greater than .80, suggesting the major source of variance is attributable to the object of measurement (e.g., person/object being observed/measured). After the G study is complete, a follow-up Decision (D) study will be conducted utilizing the data from the G study with the purpose of informing measurement improvements to decrease error as well as ideal conditions for producing dependable measurement scores (Shavelson & Webb, 1991).

**Purpose of Study**

Given the limited number of tools accessible for measuring classroom management strengths and deficiencies, it is imperative to research and enhance the reliability and dependability of available measures. An established rating training is necessary to ensure raters are producing reliable and valid results. Since behavior observation tools often have no true score index available to compare observed scores, IOA can be calculated to infer consistency and accuracy of the tool (Hintze & Matthews, 2004). This study plans to extend the results of Cassidy (2018) and Sims et al. (2021), by confirming previous findings of sufficient inter-observer agreement between the primary and secondary external raters completing the DBR-CM ER. IOA calculations will also be conducted to assess the rater agreement between the teachers completing the DBR-CM SR form and the primary external rater completing the DBR-CM ER form. An IOA score exceeding 80% is typically considered acceptable and provides
confidence that variances to the dependent variable are a result of the intervention and not error (Page & Iwata, 1986). As such, it is the goal for IOA data in this study to exceed 80%. This study hopes to improve upon previous studies (Sims et al., 2021; Cassidy, 2018), where inter-rater reliability exceeded 70% though appeared discrepant.

G Theory analyses were utilized to assess the dependability and reliability of classroom management ratings utilizing the DBR-CM ER and DBR-CM SR. Like Cassidy (2018), G theory analyses were utilized to assess the degree of error associated with possible sources of variance when rating classroom management skills: persons (p: teachers), day (d: observations), and method (m: form type). Researchers will use G Theory to assess estimations of variance in DBR-CM ratings associated with different potential sources of error and the interactions between those sources, compared to the total variance of the ratings. It is hypothesized that the greatest source of variance will be due to teachers (p) variability.
Additionally, the G study will be used to execute a D study to determine the number of ten minute observations needed to provide a reliable rating when utilizing the DBR-CM SR form or the DBR-CM ER form. The DBR-CM ER is estimated to produce dependable ratings within four observations, therefore replicating the findings from Cassidy (2018). It is hypothesized teachers will be less reliable than external raters due to biases often associated with self-report measures; however, the DBR-CM SR form is hypothesized to successfully yield adequate dependability in rating scores, and therefore can be appropriately utilized as a progress monitoring tool. As a result, it is hypothesized the DBR-CM SR form may require a slightly higher number of observations than the DBR-CM ER form to yield dependable scores. In summation, this study plans to investigate the following research questions:
(1) Is there sufficient inter-observer agreement between DBR-CM SR and DBR-CM ER forms?

(2) Is there sufficient inter-observer agreement between the primary and secondary raters completing the DBR-CM ER form?

(3) Are the DBR-CM forms dependable methods for assessing classroom management? What source accounts for the majority of rating variance?

(4) Does this study replicate previous findings of the DBR-CM ER form producing reliability standards in ratings within four, ten minute observations? How many observations, done in ten minute intervals, are required to reach reliability standards for the DBR-CM forms?
METHODS

Participants and Setting

Specific criteria for determining optimal sample sizes for G studies has not been universally endorsed (Briesch et al., 2014); however, it is often recommended that the number of data points collected is more essential than the actual number of participants. Additionally, Briesch et al. (2014) implies having fewer participants with larger number of instances is satisfactory for G Theory analyses. As a result, this study kept with consistency of previous G theory studies (Marzano, 2002; Chafouleas, Christ, Riley-Tillman, Briesch, & Chanese, 2007; Gresham, Dart, & Collins, 2017; Cassidy, 2018), and recruited a small sample with a larger number of observations.

This study recruited six, first to third grade elementary school teachers from public schools in the East Baton Rouge Parish. Due to schools closing as a result of the COVID-19 pandemic, data collection was terminated early. As a result, one teacher’s data was omitted from analyses, due to only completing 3 days of observations. First to third grade classrooms were chosen to participate in this study because recent research suggests problem behaviors should be addressed early in children’s schooling to prevent frequent occurrence or escalation of inappropriate school behaviors (Gettinger & Fischer, 2015). These grades also include consistent whole group lessons, unlike in younger years, such as kindergarten and preschool, making them ideal for needing classroom management strategies assessed.
Within a six months’ time span, teachers chose ten consecutive days where they would be observed for ten minutes each day, during the same instruction period. Due to data collection being terminated early, not all teachers were observed for the full ten days; as such, data analyses were completed based on the first seven days of data collection for each participant. Researchers recruited teachers by handing out fliers at schools within the East Baton Rouge Parish public school district. As an incentive, teachers were offered the opportunity to receive feedback on their current classroom strategies and consultation on evidence-based methods for improving strategies once their participation in the study was completed. All study procedures were implemented with each participating teacher’s classroom.

Graduate students in the school psychology doctoral program at Louisiana State University (LSU) participated in the study as trained external raters. Graduate student participants have previous training in using direct behavior rating scales and classroom management strategies. Each graduate student utilized the DBR-CM ER to rate each of the teachers, while the teachers simultaneously completed the DBR-CM SR. Each observation had one primary external rate. To calculate inter-rater reliability a portion of observations included a second external observer.

The five teachers who completed participation in the study, consisted of four 1st grade teachers and one 2nd grade teacher. All teachers were female and identified as White or Caucasian. Teachers age ranged from 23-36 years old, and their years of teaching experience ranged from 2-14 years.
Procedures

Before participant recruitment, the Institutional Review Board (IRB) at Louisiana State University approved the study proposal and administrative consent was obtained by school officials. Graduate students from Louisiana State University recruited teachers to participate in the study by handing out fliers and giving brief explanations of the study’s procedures and timeline. Informed consent was obtained from the first six teachers to inquire about participation.

A main purpose of the DBR-CM forms is to be utilized as an identifier for teachers who may benefit from additional classroom management trainings (Sims, 2016; Sims et al., 2021). Subsequently, this study examined these measures in regard to low stakes decision-making. The previous Cassidy (2018) study determined, reliable ratings for low stakes decision-making utilizing the DBR-CM ER can be established within four ten minute observations. Given the infeasibility of asking teachers to consistently rate themselves at different time increments, this study exclusively examined the number of observations related to ten minute increments. This study was expected to consist of ten consecutive observation days; however, due to schools closing as a result of the COVID-19 pandemic, data collection was terminated early for some teachers. As a result, analyses were conducted based on the first seven consecutive observation days. Due to scheduling conflicts, it was not possible for the seven days to be consecutive school days for most of the participating teachers; however, there was no larger than a two day gap between observations, and all 7 observations of a teacher were completed within three weeks (15 school days). Researchers attempted to minimize the amount of time in between observations to
minimize unobserved variables possibly influencing future data points. In total 70 data points (5 teachers x 2 methods x 7 days) were collected. This is well below the planned 120 data points (6 teachers x 2 methods x 10 days) originally intended.

When the external rater enters the classroom, they provided a copy of the self-report form to the teacher signifying the beginning of the ten minute interval. At the termination of the observation, the external rater discretely prompted the teacher to complete the self-report form, signifying the completion of the observation period. Having the presence of the external rater signify the end of the observation time period, aided in synchronizing the start and end time for both the self and external rater.

*Direct Behavior Rating-Classroom Management Forms (DBR-CM).* The DBR-CM ER (See Appendix A) and DBR-CM SR (See Appendix B) forms assess teacher classroom behavior and use a Likert-like scale rating system, ranging from 0 to 10. As mentioned previously, this scale includes four main constructs: praise, communication, enthusiasm, and rapport. Additionally, they include a single question regarding classroom structure. The back of each DBR-CM form includes detailed instructions on how to complete the form as well as detailed definitions of each behavior domain.

Classroom structure merely examines the accessibility of a classroom. This item assesses the ease to move about and see throughout the classroom for both the students and teacher. Rather than the 10-point Likert-like scale utilized for the other four domains, the classroom structure item is on a three-point interval scale: yes, somewhat, no.
Praise is operationally defined as “the use of positive praise statements in response to the behavior and performance of students in the classroom and a visibly general positive attitude towards all students” (Sims, 2016, p. 36). The next domain, communication, is operationally defined as “clearly conveying goals and expectations of a classroom and/or instructional period to students” (Sims, 2016, p. 36). The third construct, enthusiasm, is operationally defined as “the delivery of instructional content in a meaningful, memorable, and/or engaging manner” (Sims, 2016, p. 38). Lastly, rapport, is operationally defined as “the quality of the student-teacher relationship, especially that of mutual trust, emotional affinity, acceptance and positivity” (Sims, 2016, p. 37).

**DBR-CM Scoring Protocol.** The ratings of the four main domains (i.e., praise, communication, enthusiasm, and rapport) can be computed to total an overall DBR-CM score for each form. Although there are no standard scores associated with these scores, they can be used as a comparison for progress monitoring and assessment of individual teachers’ strengths and weaknesses. Ratings can be based on frequency or absolute intervals (Sims, 2014). Following the Cassidy (2018) methodology, this study based scoring on the frequency the target behavior is present within a specified timeframe (i.e., 10-minute intervals). Therefore, both the external and self-raters rated each of these four classroom management domains, from 0 to 10, based on the frequency a target behavior was present for the duration of the observation. The process for determining the frequency of a behavior was reviewed in detail in the rater training sessions required for both the external raters and teachers.
Rater Training. Rater training sessions like Sims (2020) and Cassidy (2018) studies were implemented. All external raters and teachers met with the primary researcher, where operational definitions, rating methodology of the appropriate DBR-CM form (ER or SR), and video examples and non-examples of classroom management were provided and discussed (https://dbr-cm.com/online-training). All participants were given a rater training protocol “cheat” sheet (see Appendix C). This rater training protocol “cheat” sheet was developed by Cassidy (2018) based on Sims (2016) definitions of the DBR-CM domains. This sheet included a table (see Table 1.) which graphically explains how to decipher high, medium, and low frequency of the target behaviors, as well as detailed descriptions of each frequency range. For instance, high frequency refers to the desired behaviors frequently or always occurring. If the frequency of behavior matches this definition, then it is within the 8-10 range. If the desired behavior occurs more often than the undesired behaviors, it is within the high medium frequency range of 5-7. If the desired behavior occurs less often than the undesired behaviors, it is within the low medium frequency range of 3-5. If the desired behaviors rarely or never occur, it is within the low frequency range of 0-2.

Table 1. Frequency Ranges for DBR-CM Domains

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Frequency</td>
<td>Low</td>
<td>Medium</td>
<td>High Medium</td>
<td>High Frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This training protocol sheet also included examples and non-examples of praise, communication, enthusiasm, and rapport. Participants were expected to review this sheet periodically to review rating procedures; however, participants must review the frequency break immediately prior to the observation period. All participants applied this training by watching and rating at least three example videos of teacher classroom management behavior. These ratings were compared to a master code, which was developed by the primary researcher. External raters and teachers continued this training process until 90% agreeability to the master code has been achieved.

**Design and Analysis**

*Inter Observer Agreement (IOA).* This study extended research on the reliability of the DBR-CM, by comparing the teachers’ ratings on the DBR-CM SR to that of a trained external rater on the DBR-CM ER. A recording is considered an agreement when the two observers have concurrent ratings that fall within one point, in either direction of each other (Sims, 2016). Additionally, dividing the number of domain-by-domain agreements by the total number of domains and multiplying this value by 100 will generate IOA percentages. The G study provided the variance scores among method type, however IOA analysis gives a better visual depiction of any discrepancies between the DBR-CM ER and DBR-CM SR ratings.

*Generalizability Theory Study (G Study).* This study conducted a distinct G study to evaluate the generalizability of each method of assessing classroom management. A fully crossed analytic design (p x d x m) with two facets, day (d: observations) and method (m: external and self-rater), and the object of measurement, person (p: teachers) was created. In other words, teachers were fully crossed with every day and
every rating method. Each teacher was observed for a total of ten minutes once a day for ten consecutive school days, where an external rater completed a DBR-CM ER form, while the teacher concurrently completed the DBR-CM SR form. Estimations of variance in classroom management ratings associated with each facet and the interactions between facets were compared to the total variance of ratings to examine the percentage of variance accounted for by each. A SPSS syntax specifically written for generalizability theory analyses was utilized to perform the appropriate analyses on the data collected (Mushquash & O’Conner, 2006).

Once variance components were computed, G coefficient and Phi coefficients were calculated to examine relative and absolute dependability of ratings, respectively. A follow-up decision (D) study was conducted to assess the G coefficient and Phi coefficients associated with the number of observations required to conduct a reliable rating with the DBR-CM SR and DBR-CM ER using a cutoff criterion of .80 (Briesch, et al., 2014).
RESULTS

*Missing Data.* Due to scheduling conflicts that can occur within school settings and schools closing due to social distancing policies related to the COVID-19 pandemic, some teachers were not able to be observed the full ten observation times. As such the sixth teacher was omitted from data analysis due to only receiving three days of data prior to schools closing due to the COVID-19 pandemic. Additionally, data was analyzed based on the first seven days of data collection for all teachers rather than the full ten days.

*Inter Observer Agreement (IOA).* IOA was calculated to determine the percentage of rater agreement between the primary external and self-rater, as well as between the primary external rater and the secondary external rater. IOA was determined utilizing the same definition previously stated: meeting 100% IOA was defined as the two raters having concurrent ratings that fall within one point, in either direction of each other (Sims, 2016). By dividing the number of domain-by-domain agreements by the total number of domains (i.e., five domains) and multiplying this value by 100, the following IOA percentages were generated.

The overall IOA between the primary external rater and the self-rater was calculated to be 73%. When looking at the IOA for the individual domains within the assessment, praise and enthusiasm had the weakest IOA of 60%. The IOA for the communication domain was found to be 74% and for the rapport domain IOA was found to be 71%.
IOA was also calculated to compare ratings by the primary and secondary external rater. It was originally planned to have the IOA between the primary external rater and secondary rater calculated for 30% of observations, however, due to scheduling conflicts and early termination of data collection due to COVID-19 pandemic, 28.5% of observations were conducted with a second observer. It is often suggested to collect IOA data utilizing one-third of the total data, however 20%-50% of observation sessions is thought to be acceptable (Ayres & Ledford, 2014). Again, by dividing the number of domain-by-domain agreements by the total number of domains (i.e., five domains) and multiplying this value by 100, the following IOA percentages were generated.

The overall IOA between the external and secondary raters was calculated to be 84%. When looking at the IOA for the individual domains within the assessment, praise had the weakest IOA of 70%, and fell below the acceptable 80% range. The other domains all had acceptable to good IOA estimates: classroom structure = 100%, communication = 80%, rapport = 80%, enthusiasm = 90%.

**Generalizability Theory Study (G Study).** This study conducted a distinct G study to evaluate the variance contributing to the ratings produced by the DBR-CM ER and DBR-CM SR. A fully crossed analytic design (p x d x m) with two facets, day (d: observations) and method (m: rater type), and person (p: teachers) as the object of measurement was created. In other words, teachers were fully crossed with every observation day and every rater type. Each teacher was observed for seven school days utilizing the DBR-CM ER and the DBR-CM SR during each observation. The external rater did not stay the same throughout the study. Estimations of variance in
DBR-CM ratings associated with each facet and the interactions between facets were compared to the total variance of ratings to examine the percentage of variance accounted for by each. A SPSS syntax specifically written for generalizability theory analyses was utilized to perform the appropriate analyses on the data collected (Mushquash & O’Conner, 2006).

Table 2 supplies the percentage of variance that can be attributed to various facets examined within this study. Variances can mainly be attributed to person by rater by day (66.9%). The remaining variance can be attributed to person (17.5%), person by rater (7.9%) and rater by day (7.7%). The facets of rater and day by themselves attributed to a negligible amount of variance. Although some of the variance can be attributed to person, it appears most of the variance among scores can be attributed to the combination of who is being observed, who is completing the ratings, and what day the ratings are occurring.

Table 2. Proportion of Variance for Each Facet

<table>
<thead>
<tr>
<th>Facet</th>
<th>Proportion of Variance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person (p)</td>
<td>17.5%</td>
</tr>
<tr>
<td>Method (m)</td>
<td>0*</td>
</tr>
<tr>
<td>Day (d)</td>
<td>0*</td>
</tr>
<tr>
<td>Person x Method (p x m)</td>
<td>7.9%</td>
</tr>
<tr>
<td>Person x Day (p x d)</td>
<td>0*</td>
</tr>
<tr>
<td>Method x Day (m x d)</td>
<td>7.7%</td>
</tr>
<tr>
<td>Person x Method x Day (p x m x d)</td>
<td>66.9%</td>
</tr>
</tbody>
</table>

* Attributed to a negligible amount of variance and not necessarily no variance.
Once variance components were computed, G coefficient and Phi coefficients were calculated to examine relative & absolute dependability of ratings, respectively. The G coefficient is the ratio of the universe-score variance to the expected observed-score variance and can be interpreted similarly to how a reliability coefficient in CCT is explained. The Phi coefficient measures the degree or association between two binary variables and can be interpreted like a Pearson correlation coefficient. The criterion cutoff for both the G and Phi coefficients are .80 (Briesch, Swaminathan, Welsh, & Chafouleas, 2014). The G study produced insignificant G coefficients for both the self-rater (.501) and external rater (.668) forms. It also produced insignificant Phi coefficients for both the self-rater (.486) and external-rater (.654). This suggests that assessment ratings produced insufficient relative & absolute dependability of ratings measuring classroom management skills when measured for seven observations, at ten minutes each. In other words, according to these findings, seven observations at ten minutes each is insufficient to establish a reliable estimate of a teacher's classroom management abilities via an external rater or self-rater utilizing the DBR-CM.

**Decision Study**

Since the G study suggested that the original measurement model did not produce sufficiently dependable ratings across the two methods, a decision (D) study was conducted to determine the point at which reliable ratings could be produced when manipulating the most malleable facet, Day (i.e., number of observations). Utilizing .80 as a cutoff criterion for adequate dependability (e.g., Briesch, Swaminathan, Welsh, & Chafouleas, 2014), projected G coefficients were calculated and examined for this purpose. The projected G coefficients only minimally increased when expanding the D
study out to 100 observations (See Tables 3). These results suggested in order to establish reliable ratings utilizing the DBR-CM ER, 80 observations would need to be conducted to reach a G coefficient of .80. The DBR-CM SR would require over 100 observations. These results greatly vary from previous findings (Cassidy et al., In Press) that suggested the DBR-CM ER can yield reliable estimates of classroom management skills after a trained rater observes a classroom four times, for ten minutes each observation.

Table 3. D-Study G Coefficients

<table>
<thead>
<tr>
<th># of Observations</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Rater</td>
<td>.546</td>
<td>.610</td>
<td>.634</td>
<td>.655</td>
<td>.661</td>
<td>.665</td>
<td>.668</td>
<td>.670</td>
<td>.672</td>
</tr>
<tr>
<td>External Rater</td>
<td>.706</td>
<td>.757</td>
<td>.776</td>
<td>.792</td>
<td>.796</td>
<td>.799</td>
<td>.801*</td>
<td>.803*</td>
<td>.804*</td>
</tr>
</tbody>
</table>
DISCUSSION

The current study set to further strengthen the evidence of reliability of the DBR-CM. The first part of this study implemented rating training procedures like that of Sims et al. (2021) and Cassidy (2018) in the hopes of further strengthening confidence in inter-rater reliability and therefore ensuring a degree of objectivity and consistency among observers. Specifically, this study examined:

1) Is there sufficient inter-observer agreement between DBR-CM SR and DBR-CM ER forms?

2) Is there sufficient inter-observer agreement between the primary and secondary raters completing the DBR-CM ER form?

3) Are the DBR-CM forms dependable methods for assessing classroom management? What source accounts for the majority of rating variance?

4) Does this study replicate previous findings of the DBR-CM ER form producing reliability standards in ratings within four, ten minute observations? How many observations, done in ten minute intervals, are required to reach reliability standards for the DBR-CM forms?

IOA was calculated to determine the rate of agreement among raters. IOA was first calculated to examine rater agreement between the external and self-rater. IOA between the external and self-rater ranged from 60-74% agreement. The overall IOA between the primary external rater and the self-rater was calculated to be 73%. When looking at the IOA for the individual domains within the assessment, praise and
enthusiasm had the weakest IOA of 60%. The IOA for the communication domain was found to be 74% and for the rapport domain IOA was found to be 71%. Rater agreement between teachers and external raters was poor overall, as well as across domains.

The overall IOA between the external and secondary raters was calculated to be 84%, which exceed the acceptable 80% criterion. When examining the IOA for the individual domains within the assessment, Praise had an IOA of 70%, and fell below the acceptable 80% criterion. Although Praise had the weakest IOA compared to the other domains, the 70% replicated IOA calculated for Sims et al. (69%; 2021) and Cassidy (2018). (70%; In Press). The other domains all had acceptable to good IOA estimates: Classroom Structure = 100%, Enthusiasm = 90%, Communication = 80%, and Rapport = 80%. This further supports the trainability and dependability of raters using the DBR-CM ER. However, they suggest lower agreement then the IOA calculated by Cassidy (2018), but higher agreement then the IA calculated by Sims and colleagues (2021). These discrepancies suggest unknown differences occurred either during the rater training prior to data collection or during rater maintenance strategies during data collection. Further research is needed to determine how to further improve rater agreement and decrease rater error.

Based on these findings, external raters tended to have more agreement regarding scores among each other than with teachers. Based on this data, it appears teachers struggled to adequately rate themselves similarly to the external rater. Further research is needed to determine how to improve teacher self-report. It is possible the discrepancy in ratings could be attributed to the external raters previous experience of
observation and DBR training. Additionally, although teachers were able to obtain 90% IOA during the training, the training required teachers to complete the DBR-CM ER while watching a video of another teacher. Future trainings would benefit from practicing completion of the DBR-CM SR while the teacher themselves are teaching a class.

Additionally, this study attempted to determine the dependability of the DBR-CM forms. Results suggest neither the DBR-CM ER or DBR-CM SR produced reliable ratings of the teachers classroom management skills. Although some of the variance can be attributed to person, it appears most of the variance among scores can be attributed to the combination of who is being observed, who is completing the ratings, and what day the ratings are occurring. These results varied greatly from the Cassidy (2018) findings, where the majority of the variance among DBR-CM ER scores could be attributed to the teacher being observed.

When completing the D study and expanding the number of observations out to 100 days, it was determined dependable ratings could be established with the DBR-CM ER after 80 days; however, the DBR-CM SR would require over 100 days to establish dependable ratings. Again, this contrasted greatly with the findings from Cassidy (2018) where results suggested reliable ratings could be established for the DBR-CM ER after 4 days of observations. Further research is needed to determine as to why similar results were not replicated. It is likely results were not replicated due to a lot of variety within teachers classroom management scores. See figures 2 through 6 for a representation of the teacher's total score for each observation.
Figure 2. Teacher 1 Classroom Management Performance

Figure 3. Teacher 2 Classroom Management Performance

Figure 4. Teacher 3 Classroom Management Performance
Limitations

This study’s findings must be accepted within regard to some limitations. For instance, as previously mentioned the calculated IOA between the primary external rater and self-rater was poor. Future studies would benefit from altering the training to where self-raters practice self-rating prior to data collection. Additionally, given the teachers inexperience with conducting DBR observations, a more detailed training may be warranted. In addition, IOA for the praise domain was at 70%, which is below the acceptable 80% range; however, it was similar to IOA calculated in previous studies.
(Sims et al., 2021; Cassidy, 2018). It is possible raters were unclear on expectations needed to meet criteria for this domain. Future studies should examine this domain in more detail to determine why raters struggle to agree on this domain.

Due to scheduling conflicts and the closing of schools as a result of the COVID-19 pandemic, the number of participants and days of observation were lower than originally planned. As a result, the number of data points analyzed was much lower than anticipated. It is possible the low number of data points may have negatively impacted the results of this study. Future studies should try to replicate this study with more teachers and observations days.

**Direction for Future Studies**

Future studies are necessary to examine improving rater accuracy and agreement across raters, while decreasing rater error. Additionally, this measure is thought to be a tool for school administrators to measure classroom management strategies. Given the poor agreement between teachers and external raters, further research is needed to determine the ease of training those school administrators to reliably complete observations, instead of trained graduate students with prior familiarity with DBR assessment tools.

Further, this study only utilized the DBR-CM total scores when completing the G Study. It may be advantageous to look at each domain and determine if there are specific domains that is influencing the dependability of the measure. Finally, this study only observed teachers between 1st and 2nd grade during whole group instruction. Further research is warranted to determine the generalizability of these findings to other grades.
# APPENDIX A. DIRECT BEHAVIOR RATING- CLASSROOM MANAGEMENT EXTERNAL RATER FORM

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

<table>
<thead>
<tr>
<th>Date:</th>
<th>Teacher Name:</th>
<th>Observation Start Time:</th>
<th>Instructional topic:</th>
</tr>
</thead>
<tbody>
<tr>
<td>M T W TH F</td>
<td>Observer Name:</td>
<td>End Time:</td>
<td></td>
</tr>
</tbody>
</table>

### 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 teacher to easily see one another.

<table>
<thead>
<tr>
<th>Yes</th>
<th>Somewhat</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
</tbody>
</table>

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

#### Praise
- Using positive statements or actions.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
</table>
| Low | Medium | High | Rate the level of praise used during the observation period.

#### Communication
- Clearly presenting goals and expectations.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
</table>
| Low | Medium | High | Rate the degree to which expectations were clearly communicated during the observation period.

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

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
</table>
| Low | Medium | High | Rate the degree to which the content of the observation period was delivered in an engaging manner.

#### Rapport
- The student-teacher relationship is mutually positive and accepting.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
</table>
| Low | Medium | High | Rate the degree to which the classroom climate appears to be mutually accepting, trusting, and positive.

---

Direct Behavior Rating – Classroom Management forms were created by Wesley A. Sims with support from Chris Riley-Tillman and Wendy Reinkie. Copyright © 2014 by Wesley A. Sims. All rights reserved. Permission granted to photocopy for personal and educational use, as long as the names of the creators and the full copyright notice are included in all copies.
DBR – CM ER Directions

What is DBR?
DBR is a tool that involves the brief rating of teacher classroom management behavior following a specified period of time (e.g., 25 minutes of math instruction). DBR – CM offers an efficient, flexible, defensible, and repeatable way to gather information about teacher classroom management behavior (for more information regarding DBR, see www.directbehaviorrating.org).

How to use the DBR – CM form.
Step 1: Complete observation identification information at the top of form.
Step 2: Review the definitions for each of the targeted classroom management behaviors.
Step 3: Review the directions for rating to ensure understanding of how to use the scale.
Step 4: Immediately following the observation period, rate each classroom management construct.

Example: Mrs. Jones is rating classroom management in Mrs. Pettjohn’s classroom during an Algebra lesson. In this example Mrs. Jones observes Mrs. Pettjohn asking many of her students questions and prompting them to complete example problems on the board. In addition, Mrs. Pettjohn included opportunities for her students to work through presented problems in small groups. Many of the presented problems were related to real-world scenarios. Throughout the observation, Mrs. Pettjohn’s tone was upbeat and she maintained an appropriate pace, which limited transition time between questions and activities. Based on what she observed, Mrs. Jones rated the level of enthusiasm exhibited by Mrs. Pettjohn during this period a 10.

<table>
<thead>
<tr>
<th>Enthusiasm</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruction is presented in a meaningful, memorable, and/or engaging manner</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate the level of student engagement in instruction, in conjunction with the teacher’s efforts to engage them.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Behavior Domain Descriptions

Praise is the use of positive statements or actions, including distribution of tangible reinforcers, in response to the behavior and performance of students in the classroom. In the classroom Praise looks like: Teacher uses more behavior-specific praise than general praise, uses praise contingent on expected behavior, provides three (3) or more praise statements for every reprimand, reprimands are few and when used are not harsh, teacher is more positive than negative attitude when interacting with students, provides praise at desirable rates using non-verbal interactions such as gestures, tangibles, or physical contact, and maintains an overall tone that is positive and not negative or sarcastic.

Communication refers to the clear communication of goals and expectations of an instructional period. In the classroom Communication looks like: Teacher provides clear academic and behavioral expectations to the students, explicitly states or posts instructional objectives and offers opportunity for clarifying questions, clearly presents behavioral expectations verbally and/or visually, uses an attention signal to gain attention of all students, and utilizes transition procedures that appear to be known and followed by majority of students (as evidenced by efficient classroom transitions).

Enthusiasm is the delivery of instructional content in an accurate, meaningful, memorable, and/or engaging manner; students are provided and respond to questions posed to the group and individual students frequently. Instructional material is presented using multiple modalities and is often linked to a practical, real-world example or activity. The pace and tone of instruction is upbeat and engaging, in the classroom Enthusiasm looks like: The teacher provides four (4) or more opportunities for students to respond per minute during instruction; and teacher asks many different students in the classroom at least one question during instruction. Teacher’s tone and pace of instruction are positive and upbeat, instructional content is supplemented with or related to a familiar life application; topic, or activity, and instruction incorporates alternative activities (e.g., students as teachers, group work, pair and share, current event etc.). The instructional content delivered by the teacher is accurate or correct.

Rapport is the quality of the student-teacher relationship, especially that of mutual trust, emotional affinity, acceptance, and positivity. In the classroom Rapport looks like: The general feel in the classroom is mutually warm and accepting; the teacher uses children’s names frequently; interactions between the teacher and students are visibly positive; students appear to feel comfortable approaching the teacher or asking questions; and teacher appears to feel comfortable, positive, and genuine in his/her interactions with students.

Direct Behavior Rating – Classroom Management forms were created by Wesley A. Sims with support from Chris Riley-Tillman and Wendy Reinke. Copyright © 2014 by Wesley A. Sims. All rights reserved. Permission granted to photocopy for personal and educational use, as long as the names of the creators and the full copyright notice are included in all copies.
**APPENDIX B. DIRECT BEHAVIOR RATING- CLASSROOM MANAGEMENT SELF-REPORT RATER FORM**

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

<table>
<thead>
<tr>
<th>Date:</th>
<th>Educator Name:</th>
<th>Observation Start Time:</th>
<th>Instructional topic:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M</th>
<th>T</th>
<th>W</th>
<th>TH</th>
<th>F</th>
<th>Observer Name:</th>
<th>End Time:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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.

<table>
<thead>
<tr>
<th>Yes</th>
<th>Somewhat</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Rate the level of praise used during the observation period.

**Communication**
Clearly presenting goals and expectations.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Rate the degree to which expectations were clearly communicated during the observation period.

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

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Rate the degree to which instructional content was delivered in an engaging manner.

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

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Rate the degree to which the classroom climate appeared to be mutually accepting, trusting, and positive.

Direct Behavior Rating - Classroom Management forms were created by Wesley A. Sims, Ed.D., NCSP. Copyright © 2014 by Wesley A. Sims. All rights reserved. Permission granted to photocopy for personal and educational use with inclusion of full copyright notice on all copies.
What is DBR?

DBR is a tool that involves the brief rating of educator classroom management behavior following a specified period of time (e.g., 25 minutes of math instruction). DBR-CM allows for the efficient, flexible, defensible, and repeatable assessment of educator classroom management behavior (for more information regarding DBR see www.dbrcm.com).

How to use the DBR-CM form.

Step 1: Complete observation identification information at the top of form.
Step 2: Review the definitions for each of the targeted classroom management behaviors.
Step 3: Review the directions for rating to ensure understanding of how to use the scale.
Step 4: Observe or Self-monitor.
Step 5: Immediately following the observation period, rate each classroom management construct.

Example: Mrs. Jones is rating her own classroom management during her Algebra lesson. After her lesson, Mrs. Jones reflects on her classroom management and recalls asking many of her students questions and prompting them to complete example problems on the board. In addition, she included opportunities for her students to work through presented problems in small groups. Many of the presented problems were related to real-world scenarios. Throughout the observation, Mrs. Jones judged her tone to be upbeat and she maintained an appropriate pace, which limited transition time between questions and activities. Based on her reflections, Mrs. Jones rated the level of enthusiasm in her classroom management for this period a 10.

Enthusiasm

<table>
<thead>
<tr>
<th>Instruction is presented in a meaningful, memorable, and/or engaging manner</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Rate the degree to which the content of the observation period was delivered in an engaging manner.

Behavior Domain Descriptions

Praise is the use of positive statements or actions, including distribution of tangible reinforcers, in response to the behavior and performance of students in the classroom. In the classroom, Praise looks like: Educator uses more behavior-specific praise than general praise, uses praise in response to desired student behavior, provides three (3) or more praise statements for every reprimand, reprimands are few and when used are not harsh, educator is more positive than negative when interacting with students, provides praise at desirable rates using non-verbal interactions such as gestures, tangibles, or physical contact, and maintains an overall tone that is positive and not negative or sarcastic.

Communication refers to the clear communication of goals and expectations of an instructional period. In the classroom, Communication looks like: Educator provides clear academic and behavioral expectations to the students, explicitly states or posts instructional objectives and offers opportunity for clarifying questions, clearly presents behavioral expectations verbally and/or visually, uses an attention signal to gain attention of all students, and utilizes transition procedures that appear to be known and followed by majority of students (as evidenced by efficient classroom transitions).

Enthusiasm is the delivery of instructional content in an accurate, meaningful, memorable, and/or engaging manner; students are provided and respond to questions posed to the group and individual students frequently. Instructional material is presented using multiple modalities and is often linked to a practical, real-world example or activity. The pace and tone of instruction is upbeat and engaging. In the classroom, Enthusiasm looks like: The educator provides four (4) or more opportunities for students to respond per minute during instruction; and educator asks many different students in the classroom at least one question during instruction; Educator’s tone and pace of instruction are positive and upbeat, instructional content is supplemented with or related to a familiar life application, topic, or activity, and instruction incorporates alternative activities (e.g., students as educators, group work, pair and share, current event, etc.). The instructional content delivered by the educator is accurate or correct.

Rapport is the quality of the student-educator relationship, especially that of mutual trust, emotional affinity, acceptance, and positivity. In the classroom, Rapport looks like: The general feel in the classroom is mutually warm and accepting; the educator uses children’s names frequently; interactions between the educator and students are visibly positive; the educator answers clarification questions posed by students; and the educator appears to feel comfortable, positive, and genuine in his/her interactions with students.

Direct Behavior Rating - Classroom Management forms were created by Wesley A. Sims, Ph.D., NCSP.
Copyright © 2014 by Wesley A. Sims All rights reserved. Permission granted to photocopy for personal and educational use with inclusion of full copyright notice on all copies.
**APPENDIX C. RATER TRAINING PROTOCOL**

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Frequency</td>
<td>Low Medium</td>
<td>Medium</td>
<td>High Medium</td>
<td>High Frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**High Frequency** (If matches it 9, if exceeds 10, if falls short slightly 8): The desired behaviors consistently occur

**High Medium Frequency** (If matches 6, if slightly better 7, if slightly below 5): The desired behaviors occur more often than the undesired behaviors

**Medium Frequency**: Desired behaviors occur about as often as the undesired behaviors

**Low Medium Frequency** (If matches 4, if slightly better 5, if slightly below 3): The desired behaviors occur less often than the undesired behaviors

**Low Frequency** (If matches 1, if slightly better 2, if slightly worse 0): The desired behaviors rarely occur

**More Detailed Example:**

Praise looks like:
- Positive praise statements are ideally behavior-specific, though general praise applies also
- Teachers should reward students with specific praise at a higher frequency than general praise
- Praise can be a reinforcer that is verbal, gestural, or tangible
- The student must be mindful of receiving a tangible reward for it to be considered praise and praise statements should be contingent on expected behaviors
- The number of reprimands should not exceed the 1 to 3 ratio.
- When reprimand is necessary, it should be in a calm, non-harsh voice
- The teacher should display a more positive than negative attitude and tone during student interactions

**High Frequency** (If matches it 9, if exceeds 10, if falls short slightly 8):
- Consistently praises’ all students in the class, and praise occurs throughout the entirety of the observation period
- Specific praise is greater than general praise
- Consistently at or higher than the 3 praises for every 1 reprimand ratio
- Teacher consistently displays a positive attitude and tone for the entirety of the observation
- Rarely any reprimands; Reprimands are consistently in a non-harsh, calm voice

**High Medium Frequency** (If matches 6, if slightly better 7, if slightly below 5):
- More often than not, teacher praises’ most students in the class, and praise occurs throughout most of the observation period
- Specific praise is not more frequent than general praise
- Praises are more frequent than reprimands but lower than 3 to 1 ratio (about 2:1)
- More often than not, reprimands are in non-harsh calm voice
- More often than not, teacher displays a positive attitude and tone
Low Medium Frequency (If matches 4, if slightly better 5, if slightly below 3):
- Praises’ some students in the class but not all, and praise is not consistent throughout observation period
- More often than not, praise does not occur (but s occasionally observed)
- Reprimands are more frequent or at same rate as praise
- More often than not teacher displays a negative attitude and tone

Low Frequency (If matches 1, if slightly better 2, if slightly worse 0):
- Rarely praises, Consistently/often reprimands
- Reprimands are much more frequent than praise, and are in a harsh, non-calm voice
- Teacher rarely displays a positive attitude and tone

Communication looks like:
- Teachers provide clear academic and behavioral instructions
- Explicit behavioral expectations and academic objectives can be verbally and/or visually delivered
- Time is provided for addressing questions
- Behavioral expectations and academic objectives are age appropriate and easily communicated for all students
- Teachers should communicate the behavior they expect of their students (i.e., giving corrections) rather than asking students to stop inappropriate behavior
- Students know and obey transition routines and attention signals

High Frequency (If matches it 9, if exceeds 10, if falls short slightly 8):
- Teacher consistently provides clear/explicit academic and behavioral instructions before every task and prompts reminders during tasks
- Teacher consistently provides adequate time for addressing all questions
- Instructions/expectations are consistently age appropriate, and most students consistently have access to the information (visually or verbally) (e.g., all students were attentive during teachers explanation, or can easily see and read instructions that are posted)
- Communication of inappropriate behavior is consistently phrased by telling students the expected appropriate behavior.
- All students consistently know and obey transitions routines and attention signals in a timely manner

High Medium Frequency (If matches 6, if slightly better 7, if slightly below 5):
- Teacher provides academic and behavioral instructions before most tasks and instructions are not always explicit.
- More often than not, the teacher provides time for addressing some questions
- More often than not, instructions/expectations are age appropriate, and most students have access to information (visually or verbally)
- More often than not, communication of inappropriate behavior is sometimes phrased by telling students the expected appropriate behavior, though sometimes it only consists of stating the inappropriate behavior to stop
- More often than not, students know and obey transitions routines and attention signals, and most do it in a timely manner
Low Medium Frequency (If matches 4, if slightly better 5, if slightly below 3):
- More often than not, teachers do not provide academic or behavioral instructions before tasks and instructions are not always explicit
- More often than not, teacher does not provide time for addressing questions
- More often than not, instructions/expectations are not age appropriate
- More often than not, Communication of inappropriate behaviors is not given or often only consists of listing the inappropriate behaviors and not the expected appropriate behavior
- More often than not students do not know or obey transition routines and attention signals are not often used

Low Frequency (If matches 1, if slightly better 2, if slightly worse 0):
- Teacher rarely provides academic or behavioral instructions
- Teacher rarely provides time for addressing any questions
- Instructions/expectations are rarely age appropriate
- Communication of inappropriate behavior is rarely given
- Students rarely do not know or obey transition routines and attention signals are not used

Enthusiasm looks like:
- Teachers deliver instructional content in a meaningful, memorable, and/or engaging manner
- Four or more questions should be posed per minute, with most students answering at least one question
- Minimal behavioral disruptions should be observed
- Teachers should provide accurate instruction in a positive and upbeat tone, while utilizing an appropriate pace
- Real-world examples, multiple modalities, and alternative activities (e.g., group work, current events, students teaching students, etc.) should be incorporated to supplement learned material

High Frequency (If matches it 9, if exceeds 10, if falls short slightly 8):
- Material is consistently delivered in engaging and memorable manner
- 4 or more questions are posed per minute, almost all students answer at least 1 question
- Behavioral disruptions are consistently not observed
- Instruction is consistently accurate, positive, upbeat, and utilizes an appropriate pace
- Real-world examples, multiple modalities, and alternative activities are frequently incorporated

High Medium Frequency (If matches 6, if slightly better 7, if slightly below 5):
- More often than not, the material is delivered in engaging or memorable manner
- About 4 questions or fewer are posed per minute, the majority of the students answer at least one question
- More often than not, the class is without behavioral disruptions
- More often than not, instruction is accurate, positive, upbeat, & utilizes an appropriate pace
- Real-world examples, multiple modalities, and alternative activities are incorporated sometimes

**Low Medium Frequency** (If matches 4, if slightly better 5, if slightly below 3):
- More often than not, the material is not delivered in an engaging or memorable manner
- Few questions are posed during the observation; the majority of students do not answer at least one question
- More often than not, behavioral disruptions occur
- More often than not, the instruction is not accurate, positive, upbeat or utilizing an appropriate pace
- Real-world examples, multiple modalities, and alternative activities are rarely incorporated

**Low Frequency** (If matches 1, if slightly better 2, if slightly worse 0):
- Material is rarely delivered in an engaging or memorable manner
- Questions are rarely posed during the observation, only a select few answer questions
- Behavioral disruptions are consistent
- The instruction is rarely accurate, positive, upbeat or utilizes an appropriate pace (does not have to be all 4)
- Real-world examples, multiple modalities, and alternative activities are not incorporated

Rapport looks like:
- Consists of reciprocated feelings of warmth and acceptance between the teacher and student
- Teachers frequently refer to children by their names,
- Teachers appear sincere, encouraging, and calm in their exchanges with students
- Teachers and students are both visibly comfortable and upbeat
- Students visibly seem comfortable approaching the teacher with questions or comments

**High Frequency** (If matches it 9, if exceeds 10, if falls short slightly 8):
- Reciprocated feelings of warmth and acceptance between the teacher and student are consistent
- Teacher consistently refers to children by their names
- Teachers consistently appear sincere, encouraging, and calm in their exchanges with students
- Teachers and students are consistently both visibly comfortable and upbeat
- Students consistently seem comfortable approaching the teacher with questions or comments

**High Medium Frequency** (If matches 6, if slightly better 7, if slightly below 5):
- More often than not, Reciprocated feelings of warmth and acceptance between the teacher and student are observed
- More often than not, teacher refers to students by name more often than not
- More often than not, teacher appears sincere, encouraging and calm in exchanges with students
- More often than not, teachers and students are visibly comfortable and upbeat
- More often than not, students appear comfortable to approach the teacher with questions or comments

**Low Medium Frequency** (If matches 4, if slightly better 5, if slightly below 3):
- More often than not, there are not reciprocated feelings of warmth and acceptance between the teacher and student
- More often than not, the teacher does not refer to students by names
- More often than not, the teacher does not appear sincere, encouraging, and calm with student exchanges
- More often than not, the teachers and students are not visibly comfortable and upbeat
- More often than not, students do not appear comfortable to approach the teacher with questions or comments

**Low Frequency** (If matches 1, if slightly better 2, if slightly worse 0):
- There is rarely any reciprocated feelings of warmth and acceptance between the teacher & student
- Teacher rarely refers to students by names
- The teacher rarely appears sincere, encouraging, or calm with student exchanges
- The teachers and students are rarely visibly comfortable and upbeat
- Students rarely appear comfortable to approach the teacher with questions or comments
APPENDIX D. INSTITUTIONAL REVIEW BOARD APPROVAL

ACTION ON EXEMPTION APPROVAL REQUEST

To: Kaitlin Cassidy
    Psychology

From: Dennis Landin
    Chair, Institutional Review Board

Date: August 5, 2019

Re: IRB# E11792

Title: Dependability and Reliability of the Direct Behavior Rating-Classroom Management Forms


Review Date: 8/5/2019

Approved X Disapproved ______

Approval Date: 8/5/2019 Approval Expiration Date: 8/4/2022

Exemption Category/Paragraph: 2c

Signed Consent Waived?: No

Re-review frequency: (three years unless otherwise stated)

LSU Proposal Number (if applicable):

By: Dennis Landin, Chairman

PRINCIPAL INVESTIGATOR: PLEASE READ THE FOLLOWING –

Continuing approval is CONDITIONAL on:
1. Adherence to the approved protocol, familiarity with, and adherence to the ethical standards of the Belmont Report, and LSU’s Assurance of Compliance with DHHS regulations for the protection of human subjects*
2. Prior approval of a change in protocol, including revision of the consent document or an increase in the number of subjects over that approved.
3. Obtaining renewed approval (or submission of a termination report), prior to the approval expiration date, upon request by the IRB office (irrespective of when the project actually begins), notification of project termination.
4. Retention of documentation of informed consent and study records for at least 3 years after the study ends.
5. Continuing attention to the physical and psychological well-being of informed consent of the individual participants, including notification of new information that might affect consent.
6. A prompt report to the IRB of any adverse event affecting a participant potentially arising from the study.
8. SPECIAL NOTE: When emailing more than one recipient, make sure you use bcc. Approvals will automatically be closed by the IRB on the expiration date unless the PI requests a continuation.

* All investigators and support staff have access to copies of the Belmont Report, LSU’s Assurance with DHHS, DHHS (45 CFR 46) and FDA regulations governing use of human subjects, and other relevant documents in print in this office or on our World Wide Web site at http://www.lsu.edu/irb
REFERENCES


Kunter, M. Baumert, J., & Köller, O. (2007). Effective classroom management and the development of subject-related interest. Learning and Instruction, 17, 494-509.


traditional questionnaires, and interviews. Journal of Applied Psychology, 84(5), 754-775. https://doi.org/10.1037/0021-9010.84.5.754


VITA

Kaitlin Angela Cassidy was born in Manalapan, New Jersey and attended Loyola University Maryland, in Baltimore, Maryland from August 2011 to May 2015 where she earned a Bachelor of Arts in psychology. Kaitlin earned her Master’s in psychology at Louisiana State University in 2018 and is anticipated to graduate from the School Psychology Doctoral program at Louisiana State University August 2021, under the supervision of Dr. Frank M. Gresham. Kaitlin’s clinical and research interests include behavioral interventions for students with emotional and behavioral disorders, classroom management strategies, as well as Tier 1 interventions and assessments.