The Impact of Observation Duration on the Dependability of the Direct Behavior Rating-Classroom Management External Rater Form (DBR- CM ER)

Kaitlin Angela Cassidy
Louisiana State University and Agricultural and Mechanical College, kcass9@lsu.edu

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_theses

Part of the Educational Assessment, Evaluation, and Research Commons, Elementary Education Commons, and the Elementary Education and Teaching Commons

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_theses/4662

This Thesis is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Master’s Theses by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
THE IMPACT OF OBSERVATION DURATION ON THE DEPENDABILITY OF THE DIRECT BEHAVIOR RATING- CLASSROOM MANAGEMENT EXTERNAL RATER FORM (DBR- CM ER)

A Thesis

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 Master of Arts in The Department of Psychology

by
Kaitlin Angela Cassidy
B.A., Loyola University Maryland, 2015
May 2018
# Table of Contents

Abstract ........................................................................................................................................... iii  
Introduction ..................................................................................................................................... 1  
Method .......................................................................................................................................... 11  
Results .......................................................................................................................................... 20  
Discussion ..................................................................................................................................... 25  
References ...................................................................................................................................... 28  
Appendix A: Direct Behavior Rating-Classroom Management External Rating Form .......... 34  
Appendix B: Teacher Demographic Form .................................................................................... 36  
Appendix C: Rater Training Protocol ............................................................................................ 37  
Vita ................................................................................................................................................... 42
Abstract

Classroom management skills are evidence-based strategies used to maintain a productive learning environment. The Direct Behavior Rating – Classroom Management: External Rater Form (DBR-CM ER) is an indicator for assessing classroom management. 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. Calculating inter observer agreement (IOA) assesses evidence for inter-rater reliability and IOA results were about 87%. Additionally, the study evaluated the variance contributing to the ratings produced by the DBR-CM ER. A fully crossed analytic design (p x d x t) with two facets, day (d: observations) and time (t: the time blocks), and the object of measurement, person (p: teachers) was created. Teachers (p) were identified as the greatest source of variance, with teachers (p) by day (d) the second greatest source of variance. A follow-up decision (D) study was conducted to assess the length of observations, as well as the amount of observations required to conduct a reliable rating with the DBR-CM ER. To produce a reliable and dependable rating, an observer must conduct at least four observations, where each observation is at least 10 minutes in length.
**Introduction**

Classroom management skills are evidence based, class-wide strategies used to maintain a positive, productive learning environment. These strategies include higher levels of active supervision, active student involvement, explicit performance feedback and specific and/or contingent praise (Gettinger, 1988; Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008). There is general agreement among researchers, that behavior management must be achieved in order for a productive classroom environment to occur (Doyle, 1986; Brophy, 2006; Emmer & Sabornie, 2015). Extant research has supported 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). As a result, extensive classroom management involves immediate responding to students’ current behaviors and preemptive planning for praising productive behaviors. Proactive management is purposeful in minimizing opportunities for disruptive behaviors, and consequentially, it ultimately decreases the time devoted to student discipline (Gettinger, 1988). Therefore, more time can then be devoted to class-wide academic instruction.

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, 1988; Reinke, Lewis-Palmer, & Merrell, 2008). 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). Teachers’ implementation of these strategies result in increased academic engagement, reduced classroom disruptions, and minimal individual
student disruptions (Brophy, 1985; Kounin, 1970; Simonsen et al., 2008). Conversely, classrooms where a management skill deficiency is observed, lower levels of academic performance and on-task behavior are also witnessed (Reinke et al., 2008). The absence of explicit expectations, positive teacher-student relationships, and other classroom management skills within the classroom may be catalysts for student academic failure and behavioral problems (Gottlieb & Plirstock, 2005; Split, Hughes, Wu, & Kwok, 2012). In order to correct this trend, teachers’ classroom management skills should be assessed for the purpose of strengthening those skills. Collection of assessment data (e.g. performance feedback) has a significantly positive impact on performance improvement (Fuchs, Fuchs, Hamlett, & Stecker, 1991). Improvement of teachers’ performance could reduce behavioral problems and subsequent academic failure.

**Classroom Management Support**

Although research identifies evidence-based classroom management strategies (e.g. clearly formulated rules, consistent routines, and efficient organization) as effective tools (Kunter, Baumert, & Köller, 2007; Praetius, Pauli, Reusser, Rakoczy, & Klieme, 2014), teachers typically receive little classroom management support or training regarding these strategies (Wagner et al, 2006; Stough & Montague, 2015). For instance, classroom management, as an independent course, is scarcely offered in teacher training/university programs (Brophy, 2006; Stough & Montague, 2015). Additionally, attempts to integrate classroom management skills within professional development days are often more instructional than applied (Guskey, 2000). This approach is often ineffective due to limitations in the ability to learn and master skills within a one-session workshop (Sugai et al., 2000). Sims (In Prep) suggests concurrent feedback around classroom management behavior as a possible solution. Teachers, who are recognized
with managerial deficiencies at the onset of a school year, are found to have continuing
difficulties throughout the year (Emmer & Evertson, 1981) and early identification 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 tools for
assessing classroom management easily and efficiently for the purposes of skill-level
identification and constructive feedback.

**Student Behavior Rating Scales**

Observer ratings scales are typically the preferred choice in accurately assessing
classroom instruction quality and student behavior (Praetius, et al., 2014). Traditionally,
systematic direct observation (SDO) has been considered the standard for reliable and accurate
measurement of student target behaviors (Cone, 1978; Riley-Tillman, Christ, Chafouleas, Boice-
Mallach, & Briesch, 2011). SDO is an observation technique that allows for quantifying specific
behaviors of interest in a standardized manner (Hintze, Volpe, & Shapiro, 2002; Ferguson,
Briesch, Volpe, & Daniels, 2012). School psychologists often utilize this technique to assess
student behavior for the purpose of informing decision-making processes within school settings
(Ferguson, et al., 2012). While SDO has strengths in inter-observer agreement and sensitivity to
behavior change (Briesch, Chafouleas, & Riley-Tillman, 2010), this method has been criticized
for its lack of feasibility. SDO is time sensitive and laborious, in addition to requiring extensive
rater training (Riley-Tillman, Kalberer, & Chafouleas, 2005; Riley-Tillman, et al., 2011).

A more feasible method of measurement, direct behavior rating scales (DBR), have
received recent attention for their incorporation of the positive features of SDO in addition to the
ease of traditional rating scales (Riley-Tillman, et al., 2011). DBRs employ the general outcomes measures approach in which a group of categorically related behaviors are assembled into single measurable domains (Deno, 2003). DBRs are unique in that ratings are typically conducted by those highly familiar with the measured target behaviors and occur in close temporal proximity to the target behaviors being assessed (Christ, Riley-Tillman, Chafouleas, & Boise, 2010). DBRs’ use of explicit definitions as well as, examples and nonexamples of target behaviors, allow ratings to occur quickly and repeatedly with minimal classroom interference (Christ, Riley-Tillman, & Chafoules, 2009). Further, the instrumentation and procedures of a DBR allow for immense flexibility in possible observation target behaviors; therefore, it can be utilized for various needs (Chafouleas, et al., 2013; Sims, 2016). The combination of these characteristics builds on the strengths of both SDO and behavior rating scales (Christ et al, 2009), while accounting for the weaknesses of direct observations and rating scales (Riley-Tillman et al, 2005).

**Classroom Management Assessment Tools**

Typically, DBR measures often assess target behaviors of individual students. While this is beneficial in assessing specific at-risk children, utilization of a classroom measure focusing on the management skills of the teacher, would allot school psychologists or administration the ability to assess which classrooms may be in need of implementation of evidence-based classroom management strategies. By identifying teachers or classrooms in need of management strategies, there could be an increase in effective tier I implementation and a decrease in the number of students unnecessarily referred to administration for behavioral concerns. A lack of consistent and effective strategies within the classroom may lead to student academic failures and behavioral problems, consequentially resulting in unwarranted student referrals through the
Response to Intervention (RTI) process and possibly inappropriate special education placement (Gottlieb & Plirstock, 2005). In order to ensure confidence in the RTI process or special education placement, effective universal classroom strategies must be implemented to reduce the opportunity of improper classroom management tactics reinforcing negative social or academic behaviors in students. Therefore, a tool is needed to assess which classrooms are in need of effective implementation of universal classroom management strategies, as well as to track the implementation of these strategies and their success. Due to the flexibility in the instrumentation and procedures of the DBR (Chafouleas, et al., 2013), this type of measure can be utilized for various needs, such as assessing classroom management (Sims, 2016).

**Direct Behavior Rating-Classroom Management External Rater**

The *Direct Behavior Rating – Classroom Management: External Rater Form (DBR-CM ER; Sims, 2014)* is an objective measurement of classroom management that can be utilized for the purposes discussed above (Sims, 2016). The DBR-CM ER adds to the limited number of resources utilizing a problem-solving oriented approach to rate educator performance of classroom management strategies (Sims, 2016). Initial data suggest that the measure is psychometrically sound (Sims, Riley-Tillman, Kilgus, & Reinke, In Prep; Sims, Reinke, Riley-Tillman, & Herman, In Prep.). The DBR-CM ER is composed of five subscales assessing classroom structure, praise, communication, enthusiasm, and rapport. Classroom structure simply examines the ease to move about the class given the placement of desks, materials, technology, etc. 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). Teachers’ praise has been linked to increasing student motivation, providing students with helpful and positive feedback, and establishing positive
relationships between the students and teachers (Bear, 2015); therefore, it is an essential component of classroom management. 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). Clear and explicit communication of rules and expectations is an important aspect of classroom management as it has been found to decrease the prospect of student misbehavior (Ratcliff, 2001). 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). Previous research suggests that positive student perceptions of teacher enthusiasm and interest are positively associated with students’ enjoyment and intrinsic value regarding instructional material (Keller, Goetz, Becker, Morger, & Hensley, 2014). 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). It is important to track supportive and positive relationships between students and teachers because these relationships have been found to significantly impact students social development and academic success (Hughes, Luo, Kwok, & Lyod, 2008; Hamre & Pianta, 2001; Mercer & DeRosier, 2008).

**Rater Error and Training**

Though the feasibility, repeatability, and familiarity of DBRs suggest high potential for these types of measures (Riley-Tillman et al., 2008), there is concern regarding the influence of rater error. Riley-Tillman, et al. (2009) applies the term *rater error* to describe cases where ratings tend to either over- or underestimate the true score of the target behaviors being observed via a DBR measure. Briesch et al.’s (2010) results suggested DBR rating variances were largely accounted for by rater-related effects. This is of high concern since behavioral assessments via observation should contain trustworthy scores by the raters. Even more so, The Individuals with
Disabilities Education Improvement Act of 2004 (IDEIA; 2004) requires information regarding student behavior be gathered via tools that are reliable and valid, since high stakes decisions regarding children’s education placements are often partially determined by assessment results. Error variance in relations to the DBR-CM ER has not been previously researched.

Chafouleas et al. (2013) noted the concern of rating variance within a type of DBR, the direct behavior rating - single item scales (DBR-SIS) might be impacted by the absence of systematic rater training among previous research studies regarding that measure. Further, brief DBR rater trainings involving practice and feedback have been found to improve rater accuracy (Harrison, Riley-Tillman, Chafouleas, 2014; Schlientz, Riley-Tillman, Briesch, Walcott, & Chafouleas, 2009). Therefore, one method for establishing confidence in rater accuracy of the DBR-CM ER may be conducting a routine rater training. Sims (2016) utilized a DBR 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 were discussed. Research assistants then applied the training by watching and rating example videos of teacher classroom management. Inter-observer agreement (IOA) for this study exceeded 90%; however, it is the only study that has utilized training and recorded IOA for the DBR-CM ER and replication of IOA is warranted.

**Duration of Observation Assessments**

Currently, there is no research regarding the observation duration required to yield dependable DBR-CM ER scores. Previous literature has inspected observational duration by collecting data for different observation lengths within the same interval of time. For example, McWilliam and Ware (1994) were the first to investigate student engagement by collecting individual reliability-like coefficients during the first 5 minutes, first 10 minutes, and the entire
15 minutes of the observation period. Their findings suggested as the length of an observation increases, the number of necessary observation sessions for dependable ratings of student engagement during free play decreases (e.g. twelve 15-min observations achieved a similar dependability level as seventy 5-min observations). Riley-Tillman et al. (2010) utilized a similar method for comparing observation lengths and found the longer the observation time, the more likely disruptive behaviors were overestimated when using a DBR measure. Researchers also found observation length had no significant impact on the reliability of academic engagement ratings. Conversely, Ferguson et al. (2012) found observation duration did influence the dependability of student academic engagement ratings (e.g. three 45-minute observations, four 25-minute observations, or five 15-minute observations). Ferguson et al. (2012) referred to these varying observation increments as blocks. For consistency and simplicity, these time increments will herein be referred to as time blocks. Since the DBR incorporates SDO features, Riley et al. (2011) suggests considering typical SDO periods (e.g. 5 to 20 minute time blocks) for duration of DBR assessments. Additionally, previous literature has established a threshold of five observations for lower stakes- decisions and 15 observations for higher stakes decisions to achieve reliability-like estimates on DBR assessments (Christ, et al., 2010). This study will utilize a similar methodology as the above studies and examine the necessary observation duration for assessing lower stakes-decisions with the DBR-CM ER based on these suggested SDO periods.

**Generalizability Theory**

The approach of classical test theory (CTT) is most typically used for analyzing measurements and is valuable in comprehending the degree of accuracy with which those measurements are conducted, as well as the overall strength or weakness of a measure (Briesch,
et al., 2010). However, CTT does not yield information regarding methods for reducing error or strengthening the measurement (Briesch, et al., 2010). Fortunately, the Generalizability Theory (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). G Theory has been advocated as an alternate approach for analyzing the psychometric properties of direct observation (Cone, 1978; Gresham & Carey, 1988). G Theory is a statistical framework, which examines the dependability of behavior measurements by providing information regarding 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. This partitioning permits researchers to pinpoint specific sources of measurement error that are of concern, and then evaluate the relative extent of each of those sources (Hintze & Matthews, 2004).

**Purpose of Study**

Given the minimal resources available for assessing strengths and deficiencies in teachers’ classroom behavior management skills, it is essential to research and improve reliability of available assessment tools. The first purpose of this study is to implement rater training procedures similar to that of Sims (2016) to further strengthen confidence in inter-rater reliability. Specifically, this study examines the strength of the IOA of the DBR-CM ER. Observer agreement in measures of behavior is important for ensuring a degree of objectivity and consistency among observers (Westling, Koorland, & Tait, 1981). The reliability of raters is imperative in cases where impactful decisions are being made regarding the students and teachers within a classroom. Evidence for inter-rater reliability can be assessed by calculating the IOA. It is important to have an established rating training to ensure raters are producing reliable
and valid results. Since observation measurements have no true score index available to compare observed scores to, researchers often rely on IOA to infer consistency and accuracy in observation assessments when utilizing these tools in real time (Hintze & Matthews, 2004). It is hypothesized that the results of this study will coincide with Sims (2016) results, where inter-rater reliability exceeded 90%.

The second purpose of this study is to utilize the G Theory to assess how changes in the duration and the number of observations may influence the dependability and reliability of DBR-CM ER form. Similar to Ferguson et al.’s (2012) investigation of SDO, G theory is utilized to assess the degree of error associated with possible sources of variance when using the DBR-CM ER. Explicitly, this study examines the possible proportions of variance attributed to persons (p: teachers), day (d: observations), and time (t: time blocks). G Theory is used in this study to assess estimations of variance in DBR-CM ER 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.

The G Theory outcomes can be used to execute a decision (D) study to determine the number of observations, as well as the observation length required to provide a reliable rating when utilizing the DBR-CM ER form. This is the first study utilizing G theory to examine the necessary number of observations and their required length of the DBR-CM ER. It is hypothesized that results will resemble that of previous G theory studies (e.g. McWillam and Ware, 1994; Ferguson, et al., 2012): when the length of an observation increases, the minimum number of observation sessions decrease. Since, this study is operating under the conditions of low stake decision-making, it is estimated about five observations will yield dependable scores.
Methods

Participants and Setting

This study originally included six elementary school teachers from a public school in East Baton Rouge Parish, however, due to scheduling conflicts that inhibited data collection in a timely manner, one of the second grade teachers’ participation in the study was terminated before conducting data analyses. Of the remaining teachers, three taught first grade, one taught second grade, and one taught third grade. Two of the first grade teachers were student teachers, and therefore in their first year of teaching; all other teachers in the study had three years of experience teaching at the elementary level. All participating teachers were between 22-34 years old, female, and identified as Caucasian/white.

To be included in this study, teachers had to primarily teach general education classrooms from first to third grade. These grades were chosen because recent research suggests problem behaviors should be addressed early in children’s schooling to prevent occurrence or escalation of inappropriate school behaviors (Gettinger & Fischer, 2015). Further, these grades include consistent whole group lessons, unlike in younger years, such as kindergarten and preschool where schedules are more variable depending on teaching methodology.

Researchers recruited teachers by handing out fliers at the schools. As an incentive, teachers were offered a post-participation, opportunity to receive feedback on their current classroom strategies and consultation on evidence-based methods for improving strategies. After a teacher’s participation was terminated, the primary researcher provided the teacher a summary of their results as well as psychoeducational information regarding what the DBR-CM ER is measuring and why those domains are important for maintaining a well managed classroom.
Teachers were then encouraged to seek out a consultation feedback session with researchers if they had further questions or wanted more information on methods to strengthen skills regarding classroom management strategies.

Graduate students in the school psychology doctoral program at Louisiana State University (LSU) participated in the study as data collectors. The Graduate students have previous training in systematic direct observations, behavior rating scales, and classroom management strategies. Each graduate student utilized the DBR-CM ER to rate the participating teachers. Each observation had one primary rater; for a portion of the observations, a second observer was present for calculating IOA.

Measure

**Direct Behavior Rating-Classroom Management External Rating Form (DBR-CM ER).** The DBR-CM ER (see Appendix A) assesses teacher classroom behavior as a method for identifying teachers in need of classroom management training (Sims, 2016). As mentioned previously, this scale includes four constructs: praise, communication, enthusiasm, and rapport. Additionally, the DBR-CM ER includes a single question regarding if the classroom structure allows for easy movement throughout the room, which is on a three-point interval scale: yes, somewhat, no. 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).

Similar to other DBR versions (e.g. DBR-SIS), the DBR-CM-ER uses a Likert scale rating system, ranging from 0 to 10. Initial data suggest that the measure is psychometrically sound (Sims, Riley-Tillman, Kilgus & Reinke, In Prep; Sims, Reinke, Riley-Tillman, & Herman, In Prep). For instance, the DBR-CM ER total score is significantly, positively associated with the
overall scores on the Classroom Atmosphere Measure ($r = 0.81, p < .01$) and the total with percent of positive implementation variable assessed by the Brief Classroom Interaction Observation Revised (BCIO-R; $r = .543, p < .01$; Sims, et al., In Prep). Additionally, the DBR-CM ER domains (i.e. enthusiasm, praise, communication & rapport) were found to be significantly correlated with the BCIO-R variables (i.e. rate of OTR, rate of overall praise, and rate of overall reprimands; Sims, et al., In Prep). Operational definitions and specific behavior examples and non-examples, for all of the domains, are supplied on the back of the DBR-CM ER form and are based on the criteria discussed below.

The first construct identified on the DBR-CM ER, 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). 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 in a calm, non-harsh voice (Sims, 2014). This construct also includes displaying a more positive than negative attitude and tone during student interactions (Sims, 2016).

The next 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 a majority of students knowing and obeying transition routines and attention signals (Sims, 2016; Sims, 2014).

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). 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 should be incorporated to supplement learned material (Sims, 2014).
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 (Sims, 2014).

The ratings of the four domains (i.e. praise, communication, enthusiasm, and rapport) can be added together to compute an overall DBR-CM ER score. There are no standard scores associated with these scores; however, scores can be used as a comparison for progress monitoring and assessment of individual teachers’ strengths and weaknesses. Ratings can be based on percentage of time or absolute intervals (Sims, 2014). Since this study is examining different time blocks, the ratings are based on percentage or frequency in which a target domain was observed. This method allows for the same anchoring system across differing assessment durations (Riley-Tillman et al., 2011). The rater rates the classroom on each domain, from 0 to 10, based on the frequency a target behavior was present for the duration of the time block being rated. For example, given the target behavior of praise, a rater would assign a classroom a score of 0 if praise was seen zero percent of the observation, 5 if praise was seen at a similar frequency of time as undesired behaviors, 10 if praise was seen consistently and continually throughout the entire time block.
Procedures

Before participant recruitment, the Institutional Review Board (IRB) at LSU reviewed the study methods and procedures. Additionally, administrative consent was obtained by school officials prior to participant recruitment. Graduate students from LSU recruited six teachers (only five were utilized in data collection, as previously discussed) to participate in the study by handing out fliers and giving brief explanations of the study. The first six eligible teachers to agree to participate were given an explanation of the study’s procedures and timeline. Informed consent from those six teachers was then obtained.

All study procedures were implemented with each participating teacher’s classroom. Once teachers provided consent to participate in the study, they were asked to fill out a short demographics form (See Appendix B). On this form, teachers reported their grade level, number of years teaching, race/ethnicity, contact information (i.e. room number, phone number, and email address), and a minimum of a 20-minute timeslot where whole group instruction is typically the primary teaching style, and the 10 school days they wish to participate.

Participation days were based on teachers’ availability and required 10 school days that the teacher has a whole group lesson scheduled for at least a 20-minute period. Within a two month time span, teachers’ chose 10 days where researchers observed for 20 minutes each of those days. Due to scheduling conflicts that often arise in schools (e.g. school closures, early dismissals, assemblies, class-wide testing days, etc.) it was not possible for the 10 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 ten observations of a teacher were completed within
three weeks (15 school days). Researchers attempted to minimize the amount of time in between observations in order to minimize the possibility of unobserved variables influencing future data points.

Rater Training. Training similar to Sims (2016) study was implemented to determine if similar IOA from that study (i.e., >90%) can be replicated. As per previous studies, meeting 100% IOA is defined as the two raters having concurrent ratings that fall within one point, in either direction of each other (Sims, 2016). All research assistants met with the primary researcher, where operational definitions, rating methodology of the DBR-CM ER form, and video examples and non-examples of classroom management were discussed. This was done via a training YouTube video supplied by Sims (2017). Research assistants were also supplied with a raters protocol sheet (see Appendix C), that served as a resource to remind researchers of the rating methodology during real-time observations. Research assistants then applied this training by watching and rating an example video of teacher classroom management behavior. The video was 20 minutes long and consisted of the four time blocks that would be utilized in the data collection phase. The research assistants’ ratings were compared to a master code. Similar rater training procedures were also conducted in several other studies examining behavior measures (e.g. Hintze & Matthews, 2004; Briesch et al., 2010). Research assistants completed this training until at least 90% agreement had been reached on video examples of classroom observations.

Observation schedule. A similar observation framework to previous studies examining observation duration of student behavior assessments (e.g., SDO and DBR-SIS) was utilized in this study (Ferguson et al., 2012; Riley-Tillman et al., 2011). The raters in this study filled out a
DBR-CM ER form immediately following each five minutes of the twenty-minute observation period. By the end of the 20-minute period, the rater should have had four DBR-CM ER forms completed (Table 1).

Table 1. Schedule for Completing DBR-CM ER Forms

<table>
<thead>
<tr>
<th>Time Block 1: 5 minutes</th>
<th>Time Block 2: 10 minutes</th>
<th>Time Block 3: 15 minutes</th>
<th>Time Block 4: 20 minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>At 5 minutes score DBR 1</td>
<td>At 10 minutes score DBR 2</td>
<td>At 15 minutes score DBR 3</td>
<td>At 20 minutes score DBR 4</td>
</tr>
<tr>
<td>Score DBR 1 on TB 1 only</td>
<td>Score DBR 2 on TB 1 +TB 2</td>
<td>Score DBR 3 on TB 1 + TB 2 + TB 3</td>
<td>Score DBR 4 on TB 1 + TB 2 + TB 3 + TB 4</td>
</tr>
</tbody>
</table>

To explain further, the first form was completed after the first time block (i.e., the first five minutes observed) and only those five minutes were considered when rating. At the end of the second time block the researcher filled out the form based on the frequency behaviors were observed during the total 10-minutes of the observation that had occurred thus far (which included the five minutes in the first time block). After another five minutes, a third DBR-CM ER was completed to reflect the third time block (i.e. the total 15 minutes that had been observed thus far). Finally, a last DBR-CM ER was completed immediately after the observation ended, which took into consideration the entire 20-minute observation. Researchers used a stopwatch or timer to track the time blocks. In summary, during a 20-minute period, researchers filled out four observation forms, one after every five minutes and each observation form took into consideration all of the time that had been observed up to that point.

Previous literature has established a threshold of five observations for lower stakes-decisions and fifteen observations for higher stakes decisions to achieve reliability-like estimates.
on DBR assessments (Christ, et al., 2010). A main purpose of the DBR-CM ER is to be utilized as an identifier for teachers who may benefit from additional classroom management trainings (Sims, 2016). Subsequently, this study examines this assessment tool in regards to low stakes decision-making. Since, similar research has not been conducted utilizing the DBR-CM ER form, this study consisted of 10 observation days, where four consecutive observations (i.e. 5 minute, 10 minute, 15 minute, 20 minute observations) were conducted on each day, per the five teachers (i.e. a total of 200 observations were conducted and utilized in analyses).
Results

**Missing Data.** Due to scheduling conflicts that can occur within school settings, some teachers were not able to be observed the full 10 observation times (i.e. a couple of the teachers were observed nine times). Variable mean imputation was utilized to account for this missing data. That is, that teacher mean scores for a specific time block, was used in analyses for that missing time block.

**Inter Observer Agreement (IOA).** IOA was calculated to determine the percentage of rater agreement. It was originally planned to have IOA calculated for 30% of observations, however, due to scheduling conflicts, 22% of observations were conducted with a second observer. It is most commonly suggested to collect IOA data utilizing one-third of the total data, however 20%-50% of observation sessions is acceptable (Ayres & Ledford, 2014). 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 for the DBR-CM ER was calculated to be 87%. This is slightly below the hypothesized 90% IOA, though it is still within the acceptable range for rater agreement (>80%). 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 = 93%, rapport = 86%, enthusiasm = 84%.
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. A fully crossed analytic design (p x d x t) with two facets, day (d: observations) and time (t: time blocks), and the object of measurement, person (p: teachers) was created (see Figure 1). In other words, teachers were fully crossed with every day and every time block, as each teacher was observed for ten school days and observed within each of the four time blocks. Additionally the time blocks and days were fully crossed, as each of the days consisted of the four time blocks. Each teacher was observed for a total of 20 minutes once a day for 10 school days, where a rater completed four DBR-CM ER forms, one for each time block. Though the rater did not stay the same throughout the study. Estimations of variance in DBR-CM ER 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).

Figure 1. 2-facet, fully crossed analytic design: p x d x t.
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 (49%), and to the person by day (44%). The left over variance can be attributed to person by time by day (6%), time by day (.7%) and person by time (.5%). The facets of time and day by themselves attributed to a negligible amount of variance. As a result, the majority of the variance among scores can be attributed to the actual subject being observed and the interaction between that subject and the day of the observation.

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>49</td>
</tr>
<tr>
<td>Time ((t))</td>
<td>0*</td>
</tr>
<tr>
<td>Day ((d))</td>
<td>0*</td>
</tr>
<tr>
<td>Person x Time</td>
<td>.5</td>
</tr>
<tr>
<td>Person x Day</td>
<td>44</td>
</tr>
<tr>
<td>Time x Day</td>
<td>.7</td>
</tr>
<tr>
<td>Person x Time x Day</td>
<td>6</td>
</tr>
<tr>
<td>Error</td>
<td>0*</td>
</tr>
</tbody>
</table>

* Attributed to a negligible amount of variance and not necessarily no variance.

Once variance components are 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 similar to a Pearson correlation coefficient. The G study produced a significant G coefficient of about .913 and a significant Phi coefficient which can also be rounded to .913.
This suggests that the DBR-CM ER assessment produces sufficient relative & absolute
dependability of ratings measuring classroom management skills when measured for 10
observations, at 20 minutes each.

Decision Study (D-Study). The G and Phi coefficients are utilized within the a follow-up
D-Study to determine the length of time needed, as well as the amount of observations required
to conduct a reliable rating with the DBR-CM ER within the parameters of the study (i.e. the
different lengths of time and number of observations conducted within this study), as well as,
beyond those parameters. Essentially, the D-Study is used to assess the least number of
observations and the shortest observation duration needed to produce a reliable rating of
classroom management, utilizing the DBR-CM ER.

The criterion cutoff for both the G and Phi coefficients are .80 (Briesch, Swaminathan,
Welsh, & Chafouleas, 2014). Since the G-Study suggested the original measurement model (i.e.
10 observations for 20 minutes each) more than sufficiently produced a reliable assessment for
classroom management (> .80), a D-Study was conducted to establish the most efficient number
of observations and the length those observations needed to be in order to produce sufficient
reliable estimates. The G-coefficient produced from the D-Study are seen in Table 3 and the Phi-
coefficients produced are seen in Table 4. Utilizing .80 as a cutoff criterion for adequate
dependability, for any of the measurement models (i.e. x number of observations, for y minutes
each) within Table 3 that are ≥ .80, an observer can assume reliable ratings of a teacher’s
classroom management strategies can be produced. Therefore, for the most efficient use of the
measure, a trained rater can observe a classroom four times, for 10 minutes each observation, to
sufficiently produce reliable estimates of a teachers classroom management skills.
Table 3. D-Study G Coefficients

<table>
<thead>
<tr>
<th>Number of Observations</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>Length of observations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Minutes</td>
<td>.493</td>
<td>.659</td>
<td>.742</td>
<td>.791</td>
<td>.824</td>
<td>.848</td>
<td>.866</td>
<td>.880</td>
<td>.891</td>
<td>.900</td>
</tr>
<tr>
<td>10 Minutes</td>
<td>.510</td>
<td>.674</td>
<td>.755</td>
<td>.804</td>
<td>.836</td>
<td>.859</td>
<td>.876</td>
<td>.889</td>
<td>.900</td>
<td>.909</td>
</tr>
<tr>
<td>15 Minutes</td>
<td>.515</td>
<td>.679</td>
<td>.760</td>
<td>.808</td>
<td>.840</td>
<td>.862</td>
<td>.879</td>
<td>.893</td>
<td>.903</td>
<td>.912</td>
</tr>
<tr>
<td>20 Minutes</td>
<td>.518</td>
<td>.682</td>
<td>.763</td>
<td>.810</td>
<td>.842</td>
<td>.864</td>
<td>.881</td>
<td>.894</td>
<td>.905</td>
<td>.913</td>
</tr>
</tbody>
</table>

Table 4. D-Study Phi Coefficients

<table>
<thead>
<tr>
<th>Number of Observations</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>Length of observations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Minutes</td>
<td>.490</td>
<td>.656</td>
<td>.739</td>
<td>.789</td>
<td>.823</td>
<td>.846</td>
<td>.864</td>
<td>.878</td>
<td>.890</td>
<td>.899</td>
</tr>
<tr>
<td>10 Minutes</td>
<td>.508</td>
<td>.673</td>
<td>.754</td>
<td>.803</td>
<td>.835</td>
<td>.858</td>
<td>.875</td>
<td>.889</td>
<td>.899</td>
<td>.908</td>
</tr>
<tr>
<td>15 Minutes</td>
<td>.514</td>
<td>.678</td>
<td>.759</td>
<td>.807</td>
<td>.839</td>
<td>.862</td>
<td>.879</td>
<td>.892</td>
<td>.903</td>
<td>.911</td>
</tr>
<tr>
<td>20 Minutes</td>
<td>.517</td>
<td>.681</td>
<td>.762</td>
<td>.810</td>
<td>.841</td>
<td>.864</td>
<td>.881</td>
<td>.894</td>
<td>.904</td>
<td>.913</td>
</tr>
</tbody>
</table>
Discussion

The current study set to further strengthen the evidence of reliability of the DBR-CM ER, as well as, determine the least amount of observations and the shortest observation length that is needed to yield dependable scores of a teacher’s classroom management strategies. The first part of this study implemented rating training procedures similar to that of Sims (2016) in the hopes of further strengthening confidence in inter-rater reliability and therefore ensuring a degree of objectivity and consistency among observers. IOA was calculated to determine the rate of agreement among raters. This study’s calculated IOA is above the acceptable range of 80%, however, it is slightly lower than the hypothesized estimate of 90%. This further supports the trainability and dependability of raters using the DBR-CM ER. However, further research is needed to determine how to further improve rater agreement and decrease rater error.

Additionally, this study provides further evidence that the DBR-CM ER is a dependable assessment for measuring classroom management and produces reliable ratings. For efficiency, these reliable ratings can be produced within four observations, where each lasts 10 minutes long. This further supports the efficiency and ease of using a DBR in classroom settings. Often administration takes about 20-30 minutes out of their busy schedule to assess teachers’ skills and students’ outcomes. However, this study supports that reliable ratings of classroom management can be yielded with this measure within four, 10 minute observations, saving administration time to devote to other tasks.

Moreover, the variance among scores is largely due to the object of measure (person) and the combination of the observation target and the day of the observation. This essentially supports the validity of this measure, in that it is measuring what it is claiming to measure.
(i.e. the teacher and the teacher’s relationship with her/his students) and scores typically vary as a result of the observation target themselves and/or the combination of the person and the day that person is being observed.

Limitations

This study’s findings must be accepted within regard to some limitations. For instance, as previously mentioned the calculated IOA is within the acceptable range of 80%, however, it is slightly lower than the hypothesized estimate of 90%. This could be at least partially attributed to a few different limitations. For instance, rater integrity was not assessed throughout this study. Research assistants within this study were given an original training and competency was assessed immediately following the ending of that training. However, rater integrity was not determined for those original mock ratings completed by the research assistants. Further, raters’ integrity was not assessed throughout the data collection phase. Nor was rater competency re-evaluated past the original training phase. In addition, IOA for the praise domain was at 70%, which is below the acceptable 80% range. It is possible raters were unclear on expectations needed to meet criteria for this domain. Specifically, two research assistants consistently tended to rate praise at a higher degree than other raters. More so, due to scheduling conflicts, only 22% of observations were utilized in IOA calculations. While this is in the acceptable range for calculating IOA, it is lower than the desirable 30%. It is possible that the low number of observations utilized in calculating IOA, skewed the percentages of rater agreement, either positively or negatively. The above limitations may have independently or collectively influenced rater agreement and/or rater error. Further, variable mean imputation, was utilized to account for missing data and ensure all collected data was preserved for analyses. While this is an acceptable method for account for missing data, it is possible it inflated the resulting
variances, and therefore the G and Phi coefficients. However, given the high G and Phi coefficients produced in this study, it is not suspected that this decision could have significantly inflated results.

**Direction for Future Studies**

Future studies are needed to investigate improving rater accuracy and agreement across raters, while decreasing rater error. Especially, given that higher levels of rater error is a major criticism of DBR measures. Even further, this measure is thought to be a tool for school administrations to measure classroom management strategies, further research is needed to assess the ease of training those school administrators to conduct observations, rather than trained research assistants with previous experience with similar DBR tools.

Additionally, this study only utilized the total score when conducting the G Study. It may be beneficial to look at each domain and determine if there are facets beyond person that are influencing the ratings of each specific domain. Also, more research needs to be conducted to determine why person by day influences scores, as well as, how school psychologists could assist teachers in minimizing that influence from impacting the use of good classroom management strategies. Possible theories include, that this variance in scores is the result of teachers’ interactions with students that day, events that occurred that day that impacted teachers’ behaviors, and/or events that occurred that day that impacted student behaviors. However, further research is needed to come to a clear consensus on how the interaction between person and day effect scores on the DBR-CM ER. Finally, this study only observed teachers between first and third grade during whole group instruction; more research to determine the generalizability of these findings to other grades as well as other teaching styles is warranted.
References


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


### 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></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observer Name:</th>
<th>End Time:</th>
</tr>
</thead>
<tbody>
<tr>
<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 teacher to easily see one another.

- [ ] Yes
- [ ] Somewhat
- [ ] No

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

#### Praise
Using positive statements or actions.

<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</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>
<tbody>
<tr>
<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>

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>
<tbody>
<tr>
<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>

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>
<tbody>
<tr>
<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>

Low | Medium | High
---|--------|------
Rate the degree to which the classroom climate appears to be mutually accepting, trusting, and positive.
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. Pettijohn’s classroom during an Algebra lesson. In this example Mrs. Jones observes Mrs. Pettijohn asking many of her students questions and prompting them to complete example problems on the board. In addition, Mrs. Pettijohn 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. Pettijohn’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. Pettijohn during this period a 10.

<table>
<thead>
<tr>
<th>Enthusiasm</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruction is presented in a meaningful, memorable, and/or engaging manner</td>
<td>0</td>
<td>1</td>
<td>2</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 visible positively; 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: Teacher Demographic Information

Teacher Demographic Information

Name: ______________________________
Phone number: __________________________
Email: ________________________________
Current Grade Teaching: _______________________
Room Number: _____________________________
Number of Years in Education Field: ____________
Age: __________
Sex (choose one):  ☐ Male  ☐ Female
Primary Ethnic Identity (choose one):
  ☐ African American/ Black
  ☐ Asian
  ☐ Caucasian/ White, Non-Hispanic
  ☐ Hispanic or Latino
  ☐ Native American
  ☐ Biracial/Multiracial (please specify): ____________________________
  ☐ Not Listed (please specify): ________________________________
Dates Available for Participation (ten consecutive school days):
____________________________________________________________________________
____________________________________________________________________________
____________________________________________________________________________
Best time to come in on those dates (at least 20 minute block of whole group instruction): _____
____________________________________________________________________________
____________________________________________________________________________
____________________________________________________________________________
Are you interested in learning more about how to enhance classroom management skills?
  ☐ Yes  ☐ No
Appendix C: Rater Training Protocol

20-Minute Observations

<table>
<thead>
<tr>
<th>Time Block 1: 5 minutes</th>
<th>Time Block 2: 10 minutes</th>
<th>Time Block 3: 15 minutes</th>
<th>Time Block 4: 20 minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>At 5 minutes score DBR 1</td>
<td>At 10 minutes score DBR 2</td>
<td>At 15 minutes score DBR 3</td>
<td>At 20 minutes score DBR 4</td>
</tr>
<tr>
<td>Score DBR 1 on TB 1 only</td>
<td>Score DBR 2 on TB 1 + TB 2</td>
<td>Score DBR 3 on TB 1 + TB 2 + TB 3</td>
<td>Score DBR 4 on TB 1 + TB 2 + TB 3 + TB 4</td>
</tr>
</tbody>
</table>

<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**: The desired behaviors consistently occur

**High Medium Frequency**: The desired behaviors occur more often than the undesired behaviors

**Medium Frequency** looks like (5): Desired behaviors occur about as often as the undesired behaviors

**Low Medium Frequency**: The desired behaviors occur less often than the undesired behaviors

**Low Frequency**: 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 then 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 all 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
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 is a second-year graduate student in school psychology at Louisiana State University 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 interventions for children with developmental and intellectual disabilities.