The Brief Classroom Interaction Observation–Revised: An Observation System to Inform and Increase Teacher Use of Universal Classroom Management Practices

Wendy M. Reinke, PhD[1], Melissa Stormont, PhD[1], Keith C. Herman, PhD[1], Sean Wachsmuth, PhD[2], and Lori Newcomer, PhD[1]

Abstract

Schools are increasingly using multi-tiered prevention models to address the academic and behavior needs of students. The foundation of these models is the implementation of universal, or Tier 1, practices designed to support the academic and behavioral needs of the vast majority of students. To support teachers in the use of effective Tier 1 classroom practices, researchers and practitioners need reliable and valid measures of these practices that are sensitive to change over time. The purpose of this study was to examine the reliability and validity of the *Brief Classroom Interaction Observation–Revised* (BCIO-R), which is a direct observation measure of classroom Tier 1 instructional and classroom management practices for use in elementary school classrooms. Findings indicate that the BCIO-R can be reliably implemented in the classroom context. In addition, the measure is associated with important teacher-reported constructs such as efficacy in classroom management and burnout. Furthermore, the measure is sensitive to change as indicated by demonstration of improvement in classroom management variables among teachers who received a universal classroom management intervention versus teachers who did not receive training. Having reliable and valid measures to evaluate and monitor teacher use of universal classroom practices can be useful when consulting to support teachers and improve student outcomes.

Keywords

Increasingly, schools are adopting multi-tiered prevention models to address the academic and behavior needs of students. The goal of these multi-tiered systems of support is to improve outcomes for all students through the use of evidence-based practices (Fuchs, Fuchs, & Stecker, 2010; Stormont, Reinke, Herman, & Lembke, 2012). The foundation of these models is the implementation of universal, or Tier 1, practices designed to support the academic and behavioral needs of the vast majority of students. Furthermore, Tier 1 interventions help schools determine the need for additional, more individualized supports (Tiers 2 and 3) based on a student's response, or lack of response, to Tier 1 initiatives (Stormont et al., 2012).

One important aspect of Tier 1 initiatives is the implementation of effective instructional and behavior management practices. Research has established a host of evidence-based Tier 1 classroom management practices that support student academic performance and appropriate social behavior (Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008). Despite evidence to support their use, some

teachers struggle to effectively implement universal classroom management practices and may benefit from training and consultation designed to increase their adoption and use of such practices (Reinke, Stormont, Herman, Puri, & Goel, 2011). For example, in one recent study, teachers' use of effective classroom management practices was low even though they were within a school implementing school-wide Positive Behavior Interventions and Supports (Reinke, Herman, & Stormont, 2013). Supporting teachers in the use of effective universal strategies is important given the link between teacher practices and student behaviors in the classroom (e.g., Leflot, van Lier, Onghena, & Colpin,

[1]University of Missouri, Columbia, USA
[2]Minnesota State University, Mankato, USA

Corresponding Author:
Wendy M. Reinke, Department of Educational, School, & Counseling Psychology, University of Missouri, Columbia, MO 65211, USA.
Email: reinkew@missouri.edu

2010). Research has demonstrated that teachers are responsive to training and consultation models that provide data-based feedback and that teachers often need such feedback to use specific strategies more systematically (Noell et al., 2005; Reinke, Lewis-Palmer, & Merrell, 2008).

The use of valid measures of Tier 1 classroom management practices is vital if we are to support teachers in the use of these effective practices. Despite the need, limited attention has been given to developing and validating feasible measures of Tier 1 evidence-based classroom management practices that account for the complexity of teachers needing to use varied and multiple strategies in tandem (Reddy, Fabiano, & Jimmerson, 2013). Measures that can be readily used as part of routine practices in schools could be linked to performance feedback, engage teachers in the use of Tier 1 strategies, and progress monitor teacher implementation and impact on student outcomes (Reinke, Herman, & Sprick, 2011). The *Brief Classroom Interaction Observation–Revised* (BCIO-R; Reinke & Newcomer, 2010) is an observation system that can be used to measure classroom management behaviors of teachers as well as student behaviors. The BCIO-R is an enhanced version of the tool utilized for teacher feedback in the study conducted by Reinke and colleagues in 2008. Revisions included adding several variables to the observation measure, a frequency count of harsh reprimands to distinguish them from explicit reprimands, a frequency count of teacher use of precorrective statements, and the duration of time teaching. The purpose of this study was to examine the validity and reliability of the BCIO-R for use in elementary school classrooms.

Assessing Classroom Management Practices and Needs

Valid and reliable measurement of current teacher practices and student behavior is necessary to assess and support teacher implementation of universal evidence-based interventions. Rating scales and systematic observation are two methods commonly used to measure teacher and student behavior in the classroom. Rating scales rely on student self-report and the report of others such as teachers and parents to draw conclusions about student behavior and the classroom practices of teachers. Although these reports can be valuable to recognize the initial need for classroom interventions, they lack the sensitivity and specificity of more direct measures of teacher and student behavior (Yoder & Symons, 2010). Furthermore, rating scales are not designed to be used as a repeated measure of student and teacher behavior, a feature that is necessary when implementing and monitoring the effectiveness of Tier 1 interventions.

Observation tools that assess global classroom constructs have also been developed. For instance, the *Classroom Assessment Scoring System* (CLASS; Pianta, La Paro, & Hamre, 2008) is an observational tool developed to assess classroom quality in pre-kindergarten through Grade 3 based on teacher–student interactions. The CLASS takes 30 min for observation and scoring, and is repeated up to six times over 3 hr in an effort to establish an accurate, complete picture of the classroom. Whereas the CLASS measures global constructs and takes up to 3 hr, the BCIO-R gathers the frequency and duration of specific behaviors in clearly defined contexts within a short time frame, allowing for specific data to be feasibly gathered and shared with teachers on important universal classroom practices. Data from the BCIO-R allow the researcher to develop detailed descriptions of specific behaviors and interactions in the context in which they occur. In practice, these descriptions can guide the classroom consultation process and define areas to target for improvement (Reinke, Herman, & Sprick, 2011).

Effective Classroom Management Practices

The BCIO-R was developed to gather information and support teachers' use of effective classroom management practices, monitor these practices, and evaluate the effects of these practices on student outcomes. Classroom management is an area in which teachers often feel less efficacious and request additional supports (Reinke, Stormont, et al., 2011). Thus, the BCIO-R measures teacher practices across three important classroom management domains: instructional management, promoting and responding to appropriate behavior, and discouraging and responding to inappropriate behavior. These areas are reviewed in the sections below and followed by a brief discussion of the importance of linking teacher practices with student outcomes.

Instructional Management

With regard to effective instructional classroom management practices, the BCIO-R measures teacher use of time spent on instruction and the provision of opportunities to respond. The amount of time teachers spend engaged in instruction and active supervision is related to student success (Brophy & Good, 1986). Research has shown that the most effective teachers (i.e., those with students with high levels of achievement) utilize about 50% of instructional time for instructional talk (Wallace, Anderson, Bartholomay, & Hupp, 2002). An opportunity to respond is a teacher behavior that prompts a student response, which provides immediate feedback to the teacher regarding student performance. Research has demonstrated that increasing rates of opportunities for students to respond to instruction increases student on-task behavior and decreases student disruptions (Sutherland & Wehby, 2001).

Promoting and Responding to Appropriate Behavior

Precorrection is a classroom management strategy that involves clearly stating social behavioral expectations, including routines, before students have an opportunity to make an error (Stormont & Reinke, 2009). Precorrection can be used in numerous settings to proactively promote positive behaviors in situations where students have a history of problems or before particularly difficult situations that could lead to problem behaviors (De Pry & Sugai, 2002).

Another strategy, which works very well in tandem with precorrection, is behavior-specific praise (Stormont & Reinke, 2009). Research has demonstrated that an increased teacher use of behavior-specific praise was associated with improved student time on task and fewer disruptions (Apter, Arnold, & Swinson, 2010). Behavior-specific praise is a statement that clearly indicates the desired student behavior directed to a student contingent on the student's performance of the desired behavior (Simonsen et al., 2008). Both precorrection and behavior-specific praise have been documented to reduce student problem behavior and increase appropriate behavior (Reinke, Lewis-Palmer, & Martin, 2007; Reinke et al., 2008). The BCIO-R measures teacher use of precorrective statements and both behavior-specific praise and general, or nonspecific, praise.

Responding to Inappropriate Behavior

Last, the BCIO-R measures teacher use of explicit and harsh reprimands. An explicit reprimand is a brief, contingent, and specific statement that is given when an inappropriate behavior occurs. This type of reprimand informs the student of the inappropriate behavior and provides instruction for the student on what to do in the future in a brief, concise manner. When teachers revert to making harsh or critical comments, students may actually increase disruptive behaviors in their classrooms (Van Acker, Grant, & Henry, 1996). The use of explicit reprimands following an undesired behavior, however, decreases such behavior (McAllister, Stachowiak, Baer, & Conderman, 1969). The ratio of positive to negative statements provided by a teacher can be ascertained from the data collected using the BCIO-R. Experts advise that teachers work to maintain a ratio of positive to negative interaction of at least 3:1 or 4:1 (Shores, Gunter, & Jack, 1993).

Linking Data to Student Outcomes

Gathering data on teacher practices provides important information about areas to target for improvement within a consultation relationship (Noell et al., 2005; Reinke, Herman, & Sprick, 2011). However, it is equally important to gather data on student-related behaviors to demonstrate that changes in teacher behavior do in fact correspond to positive classroom changes. Therefore, the BCIO-R also collects data on the rate of student disruptive and aggressive behaviors. Repeated measurement allows both researchers and consultants to demonstrate a functional relationship between a specific intervention and student behavior over time. Providing teachers with this information can be helpful in promoting the continued use of these universal evidence-based practices (Han & Weiss, 2005).

Current Study

The purpose of this study was to examine the reliability and validity of the BCIO-R for use in kindergarten through third-grade elementary school classrooms. We also examined whether the data gathered as part of the BCIO-R were sensitive to intervention. We expected that the BCIO-R data would be reliable when used by trained observers in elementary classrooms during instructional times (i.e., reading and math). In addition, we expected that teacher and student variables would be significantly correlated with one another in the appropriate direction. For instance, we expected that higher rates of teacher use of behavior-specific praise would be negatively correlated with student disruptive and aggressive behavior. To further determine the validity of teacher and student behavior data from the BCIO-R, we evaluated the association between the observed variables and teacher-reported efficacy in classroom management as well as teacher-reported levels of emotional exhaustion. Research indicates that self-reports of teacher efficacy are related to instructional practices (Allinder, 1994) and proactive and positive classroom management (Woolfolk, 2007). Teacher burnout is associated with increased negative interactions with students (Lamude, Scudder, & Furno-Lamude, 1992) and higher levels of student disruptive behaviors (Kokkinos, 2007). Therefore, we expected that teacher report of efficacy in classroom management would be positively associated with observed teacher use of praise and negatively associated with observed teacher use of reprimands and student disruptive behaviors. Furthermore, teacher reports of emotional exhaustion or burnout were expected to be negatively associated with observed teacher use of praise and positively associated with use of reprimands and student disruptive behaviors. Finally, we expected that data derived from the BCIO-R would be sensitive to intervention. Specifically, we expected teachers who received an intervention to demonstrate increases in observed effective classroom management practices targeted by the intervention (e.g., behavior-specific praise) at a higher rate than teachers who did not receive the intervention.

Method

The BCIO-R was utilized across 105 classrooms in nine elementary schools in an urban Midwestern school district.

Participants in this study were 105 teachers and 1,818 students in kindergarten to third grade. Of the 105 teacher participants, 97% were female. Forty-six percent of teachers reported earning a bachelor's degree as their highest degree, 47% earned a master's degree, and 7% reported earning a post-master's certificate. The racial demographics of the teachers were 22% African American, 1% Asian, 75% Caucasian, 1% Hispanic, and 1% listed as Other. Teachers taught kindergarten (27%), first (29%), second (25%), and third (20%) grades. Seventy-six percent of the teachers were between the ages of 20 and 40, whereas 24% were above the age of 41. Student participants were predominantly African American (76%), and 50% of the sample received free or reduced lunch. The number of students in the classroom ranged from 10 to 25. The number of students per classroom identified for services in special education ranged from one to five students.

Measures

Brief Classroom Interaction Observation–Revised. Independent observers conducted direct observations of teacher and student behaviors using the Multi-Option Observation System for Experimental Studies (MOOSES; Tapp, 2004) interface for hand-held computers to gather real-time data using the BCIO-R (Reinke & Newcomer, 2010). Frequency counts of teacher use of behavior-specific praise, general praise, precorrections, opportunities to respond, explicit reprimands, and harsh reprimands were gathered simultaneously during each observation. In addition, the frequencies of student disruptive behaviors and student aggressive behaviors were gathered at the same time. Last, duration of the time that the teacher was teaching was measured during the same time period. Observations were conducted across four time points during the academic year. For the purposes of this study, we used the first two data points, Time 1 before intervention and Time 2 after the initial intervention.

Because the observations were conducted as part of a larger randomized intervention trial, the fall (Time 1) observations also included individual student-level data, which were not utilized in this study. Time 1 observations were an aggregation of a series of 5-min observations, from the same observer on the same day, whereas the Time 2 observations that occurred at the start of December were 20 min in length. The range in length of the aggregated Time 1 observations was comparable, ranging from 15 min to 80 min ($M = 41.55$) across teachers. All data were converted to rate per minute, which allows observations with varying lengths to be directly compared. All observations occurred during instructional times (reading or math). Trained observers were blind to the intervention condition when observing in the classrooms across all time points. Table 1 provides the operational definitions for each observed behavior. All data were converted to rate per minute with

the exception of the time teaching, which was converted to the percent of time of the observation the teacher taught.

Teaching efficacy. The *Ohio State Teacher Efficacy Scale* (OSTES) was completed by all teachers as a measure of teacher self-efficacy (Tschannen-Moran & Hoy, 2001). The OSTES has a long and a short form to measure teacher efficacy in student engagement, instructional strategies, and classroom management. For this study, the eight items on the subscales of teacher self-efficacy of classroom management were given to teachers to answer. Teachers responded to each item by indicating their perceptions of efficacy for each item, from "nothing" (0) to "a great deal" (9). Cronbach's alphas for the subscale were .95, .96, and .95 across cohorts in this study.

Emotional exhaustion. The teacher version of the *Maslach Burnout Inventory* (MBI; Maslach, Jackson, & Leiter, 1997) is a 22-item measure assessing how frequently teachers experience feelings of burnout. Each item (e.g., "Working with students all day really is a strain for me") is measured on a 7-point scale ranging from 1 = *never* to 7 = *every day*. The MBI also contains three subscales of burnout: emotional exhaustion (nine items), depersonalization (five items), and personal accomplishment (eight items). The alpha for the total burnout measure is .87. For the purposes of this study, emotional exhaustion was the only subscale used. Example items include "I feel emotionally drained from my work" and "I feel I am working too hard on my job." The alpha for the emotional exhaustion subscale for this study is .90. The efficacy and exhaustion measures were completed by all participating teachers in October.

Procedures

All study procedures were reviewed and approved by the researcher's institution and participating school district's institutional review boards (IRBs) prior to implementation. Teachers from nine elementary schools teaching in kindergarten, first, second, and third grades were eligible for participation in the study. Interested teachers then provided informed consent to participate in the study (96% eligible consented). Next, parent consent forms were sent home to all students in the participating teachers' classrooms. Parents returned forms indicating permission for their child to be observed. Eighty-three percent of parents consented to participation in the study. Students for whom consent was not received were not included as target students during classroom observations. All teachers were randomly assigned to receive classroom management training versus not. Time 1 observations occurred before any training was provided to intervention teachers. Time 2 observations occurred after intervention teachers received the first portion of classroom

Table 1. Operational Definitions of Direct Observation Variables.

Variable	Definition
Teacher frequency codes	
Specific praise	Verbal statement or gesture that indicates approval and *names a specific behavior*
General praise	Verbal statement or gesture that indicates approval *and does not name a specific behavior*
Explicit reprimand	Verbal comments or gestures by the teacher to indicate disapproval of behavior; reprimand is concise (brief) in a normal speaking tone
Harsh reprimand	Verbal comments or gestures indicate disapproval of behavior using a voice louder than typical for the setting or harsh, critical, or sarcastic tone
Opportunity to respond	Instructional prompt (statement gesture, or visual cue) that requires immediate academic response to the teacher
Precorrective statement	Teacher provides specific statement to prompt expected student behavior(s) before the behavior occurs (e.g., before transition). The teacher must have anticipated potential for problem behavior and make the statement before problem behaviors occur. Direction prompts specific behavior expectation, not academic or content-related tasks. [Wait until the end of the statement]
Student frequency codes	
Disruptive behavior	Any statement or action by an individual student or group of students that interferes with ongoing classroom activities for the teacher and/or one or more peers
Aggressive behavior	Student is physically or verbally aggressive toward objects, peer(s), or the teacher. If physical and verbal aggressions occur simultaneously, code for both (two aggressive behaviors). You must observe the behavior to code it (do not rely on peer reporting aggression to the teacher)
Duration code: Switch following observing the behavior for 5 s	
Teaching 5-s rule	Teacher is engaged in instruction, active supervision, or is monitoring students as they work
Not teaching 5-s rule	Teacher is not engaging students and is involved in an independent task with no interaction with students

management training. A total of 52 teachers received the training and 53 did not receive the training.

Teacher training. Teachers who received the intervention in this study participated in two 6-hr workshops occurring across two consecutive days in the fall. The workshops were part of the Incredible Years Teacher Classroom Management training program (Webster-Stratton, 1994) and focused specifically on building positive relationships with students, positive attention, and use of praise. BCIO-R data were gathered at baseline prior to the first two workshops and again following Workshops 1 and 2.

Analytic Plan

Reliability checks were conducted for observations across both time points. The MOOSES program calculates reliability for each variable by determining a match between observers within a 5-s window. If a match was found, then an agreement for that variable was tallied. Variables that were not matched were tallied as disagreements. An agreement ratio was then reported for each variable (agreements divided by the sum of agreements plus disagreements). The mean percentage agreement is reported as well as the range of reliability for each variable.

Bivariate correlations among teacher and student behaviors were examined at each time point to examine how the variables were related to one another. Next, the relationship between BCIO-R data and teacher self-report of efficacy with classroom management and level of emotional exhaustion at baseline were evaluated using regression analyses. Last, regression analyses with intervention and baseline levels of the same variable as predictors were conducted using Time 2 data to determine whether BCIO-R data were sensitive to changes in teacher behavior due to receipt of training in classroom management. Because separate analyses were conducted for each variable in the regression analyses, a cutoff of $p < .01$ was used to control for the inflation of Type I error.

Results

Reliability Among Observers

Observers were trained for 2 weeks using videos and practice sessions in live classrooms to a criterion of 85% reliability with a master coder prior to conducting observations in study classrooms. Reliability checks were conducted for 29% of all observations in the field for Time 1 and 56% for Time 2. The mean percent agreement on the BCIO-R was 88% (0%–100%) for Time 1 and 90% (79%–100%) for Time 2. MOOSES utilizes a rigorous second-by-second comparison of raters to determine reliability, and an overall reliability of 80% is considered acceptable; thus, 88% and 90% are considered highly reliable (Tapp, 2004). Time 1 observations were 5-min observations later combined to be

aggregated at the classroom level; reliability for a total of 519 observations ranged from 0% to 100% across all variables at some point in the study. This shorter observation period led to instances in which the observers missed the 5-s window for agreement. In some instances, 0% agreement occurred because the behavior occurred only once during the observation and the other observer missed the behavior in the 5-s window or altogether. However, the majority of observations were reliable at 85% or higher. In instances where reliability fell below the acceptable range, observers reviewed the definitions and remained in secondary observer positions until they reached 85% reliability across variables. For the duration variable time teaching, the MOOSES calculated kappa values. Time 1 kappa was .78 (0–1.0) and that for Time 2 was .80 (.32–1.0), well above the .60 considered as acceptable reliability (Kazdin, 2011). In addition to overall reliability, averages for each teacher and student behavior were calculated. Time 1 reliability averages were 92% (0–100) for teacher precorrection, 87% (0–100) for opportunities to respond, 82% (0–100) for behavior-specific praise, 78% (0–100) for general praise, 85% (0–100) for explicit reprimands, 85% (0–100) for harsh reprimands, 85% (0–100) for student disruptive behavior, and 70% (0–100) for student aggressive behavior. Time 2 reliability averages on specific variables were 69% (0–100) for teacher precorrection, 88% (72–100) for opportunities to respond, 91% (40–100) for behavior-specific praise, 86% (56–100) for general praise, 90% (50–100) for explicit reprimands, 50% (0–100) for harsh reprimands, 90% (50–100) for student disruptive behavior, and 60% (0–100) for student aggressive behavior. Aggressive behavior agreement fell below the 80% threshold across both time points. This is likely due to the few instances of aggressive behaviors that occurred during observation times and the use of a second-by-second comparison, meaning if observers recorded the same behavior but at slightly different times (outside a 5-s window), the behavior was considered not reliably recorded. Similarly, harsh reprimands occurred infrequently and therefore if an opportunity was missed by one observer, this reduced reliability for this variable.

Intercorrelations Among Variables

Bivariate correlations among variables are presented in Table 2. Several BCIO-R teacher behaviors were significantly related, including opportunities to respond with time teaching ($r = .19, p < .05$), opportunities to respond with general praise ($r = .32, p < .01$), time teaching with precorrection ($r = .20, p < .05$), time teaching with harsh reprimands ($r = -.20, p < .05$), precorrection with behavior-specific praise ($r = .22, p < .05$), and general praise with behavior-specific praise ($r = .36, p < .01$). In addition, teacher use of reprimands was correlated with student disruptive behavior

($r = .99, p < .01$), and aggressive behavior ($r = .42, p < .01$). Furthermore, time teaching was significantly related to student aggressive behavior ($r = -.29, p < .01$). Last, student disruptive and aggressive behaviors were significantly correlated ($r = .44, p < .01$). All correlations were in the expected direction.

At Time 2, low rates of precorrection and student aggression were reported. Similarly to Time 1, several teacher behaviors were significantly related, including opportunities to respond with general praise ($r = .26, p < .05$), time teaching with behavior-specific praise ($r = .25, p < .01$), time teaching with general praise ($r = .24, p < .05$), time teaching with harsh reprimands ($r = -.52, p < .01$), precorrections with behavior-specific praise ($r = .32, p < .01$), and reprimands with harsh reprimands ($r = .29, p < .01$). All correlations were in the expected direction. In addition, teacher use of reprimands was correlated with student disruptive behavior ($r = .98, p < .01$), and aggressive behavior ($r = .34, p < .01$). Student disruptive behavior was also significantly related to teacher use of harsh reprimands ($r = .30, p < .01$). Last, student disruptive and aggressive behaviors were significantly correlated ($r = .37, p < .01$).

Association Between Teachers Reported Efficacy and Burnout

Regression analyses were conducted to determine the association between teacher-reported efficacy and burnout in the fall and direct observation of teacher and student behavior BCIO-R data gathered at the same time (see Table 3). Findings indicated that teachers who were observed to use more reprimands, and had higher rates of student disruptive behavior, self-reported lower efficacy in classroom management and higher levels of emotional exhaustion. In addition, teachers in classrooms with higher rates of student aggression reported significantly lower levels of efficacy in classroom management.

Sensitivity of BCIO-R to Intervention

Regression analyses controlling for Time 1 BCIO-R indicators were conducted to determine if intervention was associated with changes in teacher or student behavior at Time 2. Half of the teachers received two full-day workshops on the use of effective classroom management practices prior to Time 2 data collection. Baseline levels of each variable were included in the analyses to demonstrate changes in the observed behaviors over time. Findings indicated that teachers in the intervention demonstrated increases in several classroom management behaviors in comparison with teachers who did not receive the training, including precorrection ($\beta = .34, p < .001$), time teaching ($\beta = .28, p < .01$), and use of behavior-specific praise ($\beta = .38, p < .001$). No other BCIO-R indicators were significantly different between the intervention and control groups at Time 2 (see Table 4).

Table 2. Intercorrelations Among BCIO-R Variables at Time 1 and Time 2.

Variable (N = 105)	Time 1								Time 2							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
1. OTR	—								—							
2. Time teaching	**.19***	—							.12	—						
3. Precorrection	.07	**.20***	—						.07	.15	—					
4. Specific praise	-.01	-.01	**.22***	—					-.09	**.25****	**.32****	—				
5. General praise	**.32****	.08	.18	**.36****	—				**.26***	**.24***	.18	.17	—			
6. Reprimand	-.09	-.14	-.003	.04	.01	—			.05	.01	-.01	.12	.001	—		
7. Harsh reprimand	-.05	**-.20***	-.16	-.09	-.19	.17	—		.11	**-.52****	-.11	-.16	-.16	**.29****	—	
8. Disruptions	-.07	-.13	-.003	.05	.006	**.99****	**.20***	—	.06	-.01	.003	.14	.001	**.99****	**.30****	—
9. Aggression	-.15	**-.29****	.002	.04	-.06	**.42****	.11	**.44****	-.02	-.16	-.05	.006	-.02	**.34****	-.02	**.37****

Note. Values in bold indicate statistical significance. BCIO-R = *Brief Classroom Interaction Observation–Revised*; OTR = opportunity to respond.
*p < .05. **p < .01.

Table 3. Observed Variables Associated With Teacher Reported Efficacy and Emotional Exhaustion.

Observed variable	Teacher efficacy β (SE)	Emotional exhaustion β (SE)
Opportunities to respond	.18 (0.07)	−.04 (0.09)
Time teaching	.12 (0.04)	−.07 (0.05)
Precorrection	−.01 (3.24)	.06 (4.22)
Specific praise	−.02 (0.71	.12 (0.91)
General praise	.17 (0.33)	.07 (0.44)
Explicit reprimand	**−.37*** (0.20)**	**.29** (0.27)**
Harsh reprimand	−.12 (4.60)	.15 (5.93)
Student disruptions	**−.37*** (0.19)**	**.31** (0.26)**
Student aggression	**−.28** (1.69)**	.10 (2.27)

Note. Values in bold indicate statistical significance.
**p < .01. ***p < .001.

Table 4. Intervention Status Predicting Time 2 Variables While Controlling for Time 1 Variables.

Observed variable	β (SE)
Opportunities to respond	−.24 (0.49)
Time teaching	**.28** (0.60)**
Precorrection	**.34** (0.01)**
Specific praise	**.38*** (0.07)**
General praise	.09 (0.10)
Explicit reprimand	−.01 (0.07)
Harsh reprimand	−.15 (0.004)
Student disruptions	−.01 (0.08)
Student aggression	−.01 (0.02)

Note. Values in bold indicate statistical significance.
**p < .01. ***p < .001.

Discussion

The present study examined the psychometric properties of the BCIO-R as a tool for assessing and monitoring teacher classroom management practices. We found evidence to support the reliability and validity of the measure. The measure includes variables across three subdomains: instructional management, promoting and responding to appropriate behavior, and discouraging and responding to inappropriate behavior. Findings indicate that the variables are related to one another in the expected directions and several have predictive validity to teacher report of classroom management efficacy and burnout. Furthermore, the measure was able to detect changes in teacher behavior in response to receiving an intervention.

Nearly all BCIO-R items met criteria for acceptable interrater reliability across two time points, even with the rigorous real-time data collection criteria used in this study, which requires observers to agree not only on the frequency of a behavior but also the timing of it. The one exception

was the aggression frequency code, which had reliability below .80 at both time points. Given the infrequency of aggression in these K-3 classrooms and the real-time requirement, this was a stringent test of this item; for instance, if only two instances of aggression occurred during the observation and one observer only coded one instance and missed the other, or recorded it at the end of the behavior rather than start, the reliability for that variable at that time would be 50%. It is also important to note that observed aggression on the BCIO-R had a significant negative association with teacher-reported self-efficacy suggesting that these observations were assessing a relevant aspect of the classroom environment; if anything, we expect this significant relationship to be attenuated by the lower reliability of the aggression scale, so this is a promising finding. Future use of the aggression variable in settings with higher rates of aggression (e.g., secondary school settings) may allow for higher rater agreement.

Interrelationships among the items provided evidence of convergent and discriminant validity. Items within domains had small and medium relationships among each other suggesting that they were measuring similar but distinct constructs. For example, general praise, specific praise, and precorrections were all positively correlated demonstrating convergent validity in that these variables evaluate aspects of a broader construct, proactive classroom management strategies. Furthermore, both types of reprimands were negatively correlated with these proactive strategies demonstrating discriminant validity.

Although the variables assessed by the BCIO-R can all stand alone in giving teachers feedback on their performance, it may also be useful to combine some of the BCIO-R variables to evaluate teacher change over time. For instance, in a recent study, we created a composite variable that indicated the percentage of proactive strategies (e.g., praise and precorrections) versus reprimands, or reactive strategies, used by teachers (see Reinke, Stormont, Herman, & Newcomer, 2014). This composite variable allows for researchers to determine if teachers utilized more proactive strategies in proportion to reactive strategies over time.

Teacher-reported measures provided evidence in support of the predictive validity of several of the BCIO-R variables. Teacher efficacy was negatively related to observed rates of reprimands, student disruptions, and student aggression. Moreover, teacher-reported emotional exhaustion was positively related to observed rates of reprimands and student disruptions. This aligns with past research that has demonstrated that classroom disruptive behaviors were associated with emotional exhaustion burnout (Hastings & Bham, 2003), and teachers with lower efficacy used less effective practices (Woolfolk, 2007).

These findings also have implications for identifying teachers in need of support. Direct observations on the BCIO-R suggest that teachers with observed high rates of

reprimands and whose classrooms have high rates of student disruptions and aggression may be experiencing risk of burnout. It is likely that these factors influence each other transactionally; that is, teachers with high rates of stress are more likely to use reactive strategies and classrooms with high rates of disruptions are likely to increase teacher stress. Thus, providing teacher supports in the form of not only coping and stress management but also reducing the reprimand–disruption cycle by using more effective classroom management practices would likely yield benefits for both teachers and students.

Finally, several variables measured by the BCIO-R were sensitive to intervention. Time teaching was significantly higher at Time 2 for teachers who received an intervention compared with teachers who did not. Time teaching is a variable with strong face validity in school settings. Teachers who spend more time teaching are likely to have students who demonstrate increased academic performance. Future research should determine if improvements in time teaching as assessed by the BCIO-R is positively associated with student academic gains.

In addition, precorrections and specific praise showed intervention effects with teachers who received the intervention using these strategies at higher rates than control teachers. The training that teachers received focused predominantly on the use of proactive classroom management strategies, and therefore it was expected that teachers would increase their use of observed praise and precorrective statements following such training. Interestingly, the BCIO-R did not distinguish intervention classrooms on student disruptive or aggression behaviors. The teachers in this study were not selected because they needed help with classroom management necessarily. All teachers in the K-3 classrooms were eligible to be randomly assigned to condition. Therefore, the lack of findings for decreasing disruptive or aggressive behavior could be that the classrooms simply had very few of these student behaviors. Therefore, future research may target classrooms of teachers who are struggling with classroom management and who are experiencing high levels of student disruptive behavior to evaluate the sensitivity of the BCIO-R to intervention on student outcomes. Also, longer term follow-up from the time of intervention may be useful. Future research that focuses on opportunities to respond should be conducted to determine if this variable is as sensitive to change.

Findings from this study have implications for improving classroom management practices in elementary schools. The BCIO-R is a reliable and valid indicator of critical aspects of classroom management. The tool can be used to study both classroom management processes and how they change over time, naturally or after intervention. Importantly, the tool can also be used in practice to support teachers in developing more effective practices. By monitoring practices over time and giving feedback to teachers about their practices, the BCIO-R can be used as a progress monitoring and performance feedback tool. Abundant research supports the importance of performance feedback in coaching teachers to develop new skills (Noell et al., 2005; Reinke et al., 2013). Therefore, having a tool such as the BCIO-R to guide the provision of performance feedback within a consultation relationship should lead to positive outcomes for teachers and students. Consultation models, such as the Classroom Check-Up (Reinke, Herman, & Sprick, 2011), already have performance feedback embedded into the model as a method for supporting teacher use of classroom management practices. Additional research on the use of the BCIO-R within these consultation models could help to inform how best to support teachers in using universal practices effectively.

The BCIO-R code for this study was implemented as part of a real-time data collection process using MOOSES. Using MOOSES as a platform for the BCIO-R is optimal, particularly in the context of a research study. However, the BCIO-R can also be implemented in a paper-and-pencil format that would likely make it more feasible for practitioners. In the past, consultants have used the BCIO-R in a paper-and-pencil format that did not include the duration variable of time teaching to develop intervention plans and monitor progress with teachers (see Reinke, Herman, & Sprick, 2011). School psychologists or other practitioners could use a paper-and-pencil frequency count version of the BCIO-R to support teachers struggling with classroom management. In addition, administrators could use a similar version of the BCIO-R when conducting walk-throughs and providing feedback to teachers (see Sprick, Knight, Reinke, Skyles, & Barnes, 2010). Furthermore, teachers could use the BCIO-R paper-and-pencil version to observe one another and provide peer coaching to help improve classroom management and sustain behavior change over time. Research focused on the use of the BCIO-R with applied practitioners would be helpful in understanding how a measure focused on direct observation of universal practices can be feasibly and effectively implemented.

Limitations

It is important to note that the study is not without limitations. The study occurred in the context of a larger randomized trial. It is not known how the findings will generalize outside the context of the study. Also, it is ideal to conduct reliability checks on 30% of observations, and thus, the 29% conducted for Time Point 1 is slightly less than optimal. Time 1 and Time 2 observations were conducted slightly differently, but converted to a common metric for comparison. However, this difference between how the observations were conducted may have affected findings. Future studies that use only the 20-min format of the BCIO-R should be conducted. Of particular importance,

further research is needed to assess the reliability and validity of the BCIO-R in practice settings by school psychologists and other school professionals. Future research should be conducted to determine the feasibility of the BCIO-R in applied practice settings. In addition, though most relationships between variables and predictors were in the expected direction, some hypotheses were not supported. Future research with a larger sample of teachers may be helpful in identifying significant relationships. However, all of the variables measured within the BCIO-R have high face validity and utility. The variables are all critical indicators of effective classroom management practices. When these variables are assessed reliably in the context of classrooms, each provides important information toward supporting teachers who are struggling with universal classroom management practices. Last, only kindergarten through third-grade elementary teachers were included in this study; thus, the findings will not likely translate to secondary education classrooms. Future research should implement the BCIO-R across middle school and upper elementary classrooms to determine how the rates of teacher and student behaviors may vary in these contexts. In addition, systematic studies of the BCIO-R to examine the psychometric properties of this tool are warranted. For instance, future research might be conducted to further strengthen the confidence in the use of the tool including using another well-established observational measure for convergent validity.

Conclusion

Multi-level prevention models within schools can be helpful toward improving academic and social behaviors of students (Stormont et al., 2012). However, some teachers require additional supports to effectively implement universal prevention practices within their classrooms (Reinke, Herman, & Sprick, 2011). Having reliable and valid measures to evaluate and monitor teacher practices in the use of universal classroom practices can be useful when consulting to support teachers. The BCIO-R was designed to assess teacher behaviors so that consultants can make targeted suggestions regarding potential evidence-based interventions and provide support for the implementation of the interventions. The BCIO-R is a universal classroom measure that can be used to provide teachers with performance feedback on critical classroom management indicators, which can lead to improved practices for teachers, fewer disruptive behaviors, and improved academic outcomes for students. When universal practices are in place, students who are in need of additional supports, because they do not respond to these universal supports, can be more accurately identified and provided higher tiered interventions. Thus, using tools like the BCIO-R to support teachers in using universal prevention practices is paramount toward maximizing school resources, which are often limited.

Authors' Note

The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

Allinder, R. M. (1994). The relationship between efficacy and instructional practices of special education teachers and consultants. *Teacher Education and Special Education*, *17*, 86–95.

Apter, B., Arnold, C., & Swinson, J. (2010). A mass observation study of student and teacher behaviour in British primary classrooms. *Educational Psychology in Practice*, *26*, 151–171.

Brophy, J. E., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–375). New York, NY: Macmillan.

De Pry, R. L., & Sugai, G. (2002). The effect of active supervision and pre-correction on minor behavioral incidents in a sixth grade general education classroom. *Journal of Behavioral Education*, *11*, 255–267. doi:1053-0819/02/1200-0255/0

Fuchs, D., Fuchs, L., & Stecker, P. (2010). The "blurring" of special education in a new continuum of general education placements and services. *Exceptional Children*, *76*, 301–323.

Han, S. S., & Weiss, B. (2005). Sustainability of teacher implementation of school-based mental health programs. *Journal of Abnormal Child Psychology*, *33*, 665–679.

Hastings, R. P., & Bham, M. S. (2003). The relationship between student behaviour patterns and teacher burnout. *School Psychology International*, *24*, 115–127.

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York, NY: Oxford University Press.

Kokkinos, C. M. (2007). Job stressors, personality and burnout in primary school teachers. *British Journal of Educational Psychology*, *77*, 229–243.

Lamude, K. G., Scudder, J., & Furno-Lamude, D. (1992). The relationship of student resistance strategies in the classroom to teacher burnout and teacher Type-A behavior. *Journal of Social Behavior and Personality*, *7*, 597–610.

Leflot, G., van Lier, P. A. C., Onghena, P., & Colpin, H. (2010). The role of teacher behavior management in the development of disruptive behaviors: An intervention study with the good behavior game. *Journal of Abnormal Child Psychology*, *38*, 869–882.

Maslach, C., Jackson, S. E., & Leiter, M. P. (1997). Maslach Burnout Inventory. In C. P. Zalaquett & R. J. Wood (Eds.)., *Evaluating stress: A book of resources* (3rd ed., pp. 191–218). Lanham, MD: Scarecrow Education.

McAllister, L. W., Stachowiak, J. G., Baer, D. M., & Conderman, L. (1969). The application of operant conditioning techniques in a secondary school classroom. *Journal of Applied Behavior Analysis*, 2, 277–285.

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

Pianta, R., La Paro, K., & Hamre, B. (2008). *Classroom Assessment Scoring System: Manual K-3*. Baltimore, MD: Brooks Publishing.

Reddy, L. A., Fabiano, G., & Jimmerson, S. (2013). Assessment of general education teachers tier 1 classroom practices: Contemporary science, practice and policy. *School Psychology Quarterly*, 28, 273–276.

Reinke, W. M., Herman, K. C., & Sprick, R. (2011). *Motivational interviewing for effective classroom management: The classroom check-up*. New York, NY: Guilford Press.

Reinke, W. M., Herman, K. C., & Stormont, M. (2013). Classroom level positive behavior supports in schools implementing SW-PBIS: Identifying areas for enhancement. *Journal of Positive Behavior Interventions*, 15, 51–60.

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

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

Reinke, W. M., Stormont, M., Herman, K. C., & Newcomer, L. (2014). Using coaching to support teacher implementation of classroom-based interventions. *Journal of Behavioral Education*, 23, 150–167.

Reinke, W. M., Stormont, M., Herman, K. C., Puri, R., & Goel, N. (2011). Supporting children's mental health in schools: Teacher perceptions of needs, roles, and barriers. *School Psychology Quarterly*, 26, 1–13.

Shores, R. E., Gunter, P. L., & Jack, S. L. (1993). Classroom management strategies: Are they setting events for coercion? *Behavioral Disorders*, 18, 92–102.

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education & Treatment of Children*, 31, 351–380. doi:10.1353/etc.0.0007

Sprick, R., Knight, J., Reinke, W. M., Skyles, T., & Barnes, L. (2010). *Coaching classroom management: Strategies and tools for administrators and coaches* (2nd ed.). Eugene, OR: Pacific Northwest Publishing.

Stormont, M., & Reinke, W. M. (2009). The importance of pre-correction and behavior specific praise strategies. *Beyond Behavior*, 18, 26–32.

Stormont, M., Reinke, W. M., Herman, K. C., & Lembke, E. (2012). *Academic and behavior supports for at-risk students: Tier 2 interventions*. New York, NY: Guilford Press.

Sutherland, K. S., & Wehby, J. H. (2001). The effect of self-evaluation on teaching behavior in classrooms for students with emotional and behavioral disorders. *The Journal of Special Education*, 35, 161–171.

Tapp, J. (2004). *Multi-Option Observation System for Experimental Studies (MOOSES)*. Retrieved from http://kc.vanderbilt.edu/mooses/mooses.html

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, 17, 783–805.

Van Acker, R., Grant, S. H., & Henry, D. (1996). Teacher and student behavior as a function of risk for aggression. *Education & Treatment of Children*, 19, 316–334.

Wallace, T., Anderson, A. R., Bartholomay, T., & Hupp, S. (2002). An ecobehavioral examination of high school classrooms that include students with disabilities. *Exceptional Children*, 68, 345–359.

Webster-Stratton, C. (1994). *The Incredible Years teacher training series*. Seattle, WA: The Incredible Years.

Woolfolk, A. (2007). *Educational psychology* (10th ed.). Boston, MA: Allyn & Bacon.

Yoder, P., & Symons, F. (2010). *Observational measurement of behavior*. New York, NY: Springer.