Hammill Institute on Disabilities

Journal of Positive Behavior Interventions
2021, Vol. 23(3) 163–173

DOI: 10.1177/1098300720935101
jpbi.sagepub.com
SAGE

"Stop Doing That!": Effects of Teacher Reprimands on Student Disruptive Behavior and Engagement

Paul Caldarella, PhD[1], Ross A. A. Larsen, PhD[1], Leslie Williams, EdS[1], Howard P. Wills, PhD[2], and Joseph H. Wehby, PhD[3]

Abstract

Many teachers resort to using reprimands in attempts to stop disruptive student behavior, particularly by students with emotional or behavioral problems, although this may not be effective. This study examined short-term longitudinal data on teacher reprimands of 149 teachers in 19 different elementary schools across three states, as well as disruptive behavior and classroom engagement of 311 students considered at risk for emotional and behavioral disorders. A cross-lag analysis showed that teacher reprimands did not decrease students' future disruptive behavior or increase their engagement or vice versa. While teacher reprimands may suppress misbehavior momentarily, they do not appear to be effective in decreasing students' disruptive behavior or increasing their engagement over time. Limitations and implications are discussed.

Keywords

teacher reprimands, student disruptions, student engagement, elementary school

Students with emotional and behavioral disorders (EBD) experience many challenges in school. The literature has identified common characteristics of students with or at risk for EBD, including aggression, attention and academic problems, antisocial behavior, low classroom engagement, high rates of disruptive behaviors, and mental health challenges (Caldarella et al., 2019; Conley et al., 2014; Salle et al., 2018). These characteristics, in addition to disrespect and hyperactivity, were reported by elementary teachers as common behavior problems observed in their classrooms (Conley et al., 2014). Teachers also report that students at risk for EBD share many of these characteristics, as well as frequent noncompliance with teacher directions (Hecker et al., 2014).

The ways in which teachers and students interact can affect outcomes for students with EBD. For example, teachers' struggles to manage classroom behavior may be related to these students' negative academic outcomes (Caldarella et al., 2019). Teachers report feeling unprepared to implement effective strategies to serve at-risk students, particularly those with emotional or behavioral problems (Chafouleas et al., 2010; Reinke et al., 2011). Student–teacher relationships can also be more challenging for students with EBD, as they often refuse to follow directions, engage in defiant/non-compliant behaviors, or ignore their teachers in a passive-aggressive manner (Hecker et al., 2014). However, there can be positive outcomes if the teacher–student interactions are positive. For example, teacher–student relationships can be improved by simply giving students a personalized greeting each day (Allday & Pakurar, 2007). Furthermore, with such personalized greetings, teachers have been able to increase the on-task behavior, or engagement, and decrease disruptions of students in their classrooms (Cook et al., 2018). In addition, teachers who deliver low rates of negative feedback (e.g., reprimands) and high rates of positive feedback (e.g., praise) may be particularly effective with students with EBD when providing multiple teaching and learning opportunities that enhance students' engagement (Rathel et al., 2014).

Reprimands

Reprimands, often intended as a form of punishment, are statements meant to correct misbehavior (Allday et al., 2012) and to decrease the probability or frequency of the behavior they follow. Reprimands have also been referred

[1]Brigham Young University, Provo, UT, USA
[2]The University of Kansas, Kansas City, USA
[3]Vanderbilt University, Nashville, TN, USA

Corresponding Author:
Paul Caldarella, Center for the Improvement of Teacher Education and Schooling, Brigham Young University, 149 MCKB, Provo, UT 84602, USA.
Email: paul_caldarella@byu.edu

Action Editor: Maureen Conroy

to as *negative communication* (Rathel et al., 2008) or *contingent punishment* (Merrell et al., 2012). Reprimands are noted as being less effective than positive behavioral classroom management strategies (Reinke et al., 2013). Students with or at risk for EBD appear to become more engaged and less disruptive when they receive less frequent teacher reprimands and more frequent teacher praise, responding more powerfully to these strategies than their typically developing peers (Downs et al., 2019).

In elementary school classrooms, reprimand rates have been observed to be higher than praise rates (Reinke et al., 2013; Van Acker et al., 1996) and tend to increase as grade level progresses (Reddy et al., 2013; White, 1975). However, in one study, while students at risk for EBD received approximately equal rates of praise as their peers, they received a significantly greater number of reprimands, thus creating an unfavorable praise to reprimand ratio (1:1.92 for at-risk students compared with 1:1.43 for peers; Caldarella et al., 2020). In a study by Downs et al. (2019), students with emotional and behavior problems received reprimands at a rate of 0.10 per minute (SD = 0.07) compared with their peer comparisons who received 0.04 reprimands per minute (SD = 0.05).

Teachers often respond to disruptive student behavior by increasing their use of reprimands (Hollingshead et al., 2016), which may temporarily stop misbehavior (Alber & Heward, 2000). If misbehavior stops, the teacher's behavior (i.e., reprimanding) is reinforced and teachers continue to utilize reprimands (Shores et al., 1993). Although reprimanding may be less effective in the long term, teachers may continue to engage in reprimands for a variety of reasons including lack of training/professional development, constraints on time, or burnout (Jennings & Greenberg, 2009) in addition to personal reinforcement.

From an operant framework (Skinner, 1938), the application of an unpleasant stimulus results in a decrease in the behavior it follows. Thus, reprimands would be predicted to decrease student disruptive behaviors, if applied consistently and if reprimands are experienced as an unpleasant stimulus. The antecedent of student misbehavior may result in the behavior of teacher reprimands, the consequence of which may be decreased student misbehavior, thus reinforcing teacher reprimands. There is however the possibility that students may engage in disruptive behavior to obtain teacher attention, be it positive or negative attention, resulting in reinforcement of disruptive behavior (Martens & Ardoin, 2010). Another consequence of teacher reprimands could be to decrease student disengagement, potentially resulting in increased student engagement. These hypotheses can be tested using a cross-lag model (Kenny, 1975) where teacher reprimands, student disruptive behavior, and student engagement at time t are regressed on those same measures at time $t - 1$, allowing for any causal impacts of teacher reprimands, student disruptive behavior, and student engagement on each other to be isolated.

Research shows mixed effects of teacher reprimands on student behavior (see overview by Gable et al., 2009). For example, a small study during a summer school with 18 students with behavioral and academic difficulties (Sherrill et al., 1996) found that the more consistently the teacher reprimanded, the less disruptive students were (specifically regarding callouts). This study showed that consistent reprimanding did not negatively affect overall participation or hand-raising. However, other researchers have found that teacher reprimands can cause students to engage in verbal or physical aggression (Van Acker et al., 1996) and escape-motivated behaviors (Shores et al., 1993), such as acting out to be removed from the classroom during math instruction due to low math ability. More recent findings have also suggested potentially negative outcomes associated with the use of reprimands including positive correlations with students' problem behavior (Downs et al., 2019; Kodak et al., 2007) and teachers' emotional exhaustion (Reinke et al., 2013), and negative correlations with students' on-task behavior or engagement (McComas et al., 2017). However, there is a lack of longitudinal studies examining the impact of teacher reprimands on student disruptions and engagement. Given some differences between past research findings, as well as the lack of predictive studies using longitudinal data, further evaluation of the effects of reprimands on students' classroom behavior is warranted to help fill a gap in the literature.

Study Purpose

Although reprimands are used more frequently in classrooms than praise (Reinke et al., 2013), recent research on the effectiveness of reprimands is inadequate. Although correlational studies have linked reprimands to variables such as disruptive behavior and reduced student engagement (Downs et al., 2019; McComas et al., 2017), limited studies have examined longitudinal data and causation. In this study, we used a cross-lag analysis to examine the effects of teacher reprimands on future student behavior while controlling for teacher praise, given the potential association between teacher praise and student behavior. Four specific questions guided this research:

1. Do teacher reprimands decrease future disruptive behavior of students at risk for EBD after accounting for past teacher reprimands and past student disruptive behavior?
2. Does student disruptive behavior increase future teacher reprimands toward students at risk for EBD after accounting for past teacher reprimands and past student disruptive behavior?

3. Do teacher reprimands increase future engagement of students at risk for EBD after accounting for past teacher reprimands and past student engagement?
4. Does a lack of student engagement increase future teacher reprimands toward students at risk for EBD after accounting for past teacher reprimands and past student engagement?

Method

Settings and Participants

Data for this study were collected from 19 elementary schools in Missouri, Tennessee, and Utah as part of a 4-year randomized control trial (RCT) of a proactive classroom management program called Class-Wide Function-related Intervention Teams (CW-FIT; Wills et al., 2010). Teachers were randomly selected to participate in the CW-FIT program (see Wills, Wehby, et al., 2018, and Wills, Kamps, et al., 2018, for more information). However, treatment condition assignments were not a factor in this study, as only baseline data were considered. Participants included 311 students identified as at risk for EBD (see the "Procedures" section) and 149 teachers. There were approximately two at-risk students per class. See Table 1 for participant demographic information. The average school free or reduced-price lunch percentage was 74.57% (SD = 19.44).

Table 1. Participant Demographic Data.

Variable	At-risk students (n = 311) Frequency	%	Teachers (n = 149) Frequency	%
Grade level				
Kindergarten	55	17.68	27	18.12
First grade	58	18.65	29	19.46
Second grade	50	16.08	22	14.77
Third grade	66	21.22	31	20.81
Fourth grade	35	11.25	20	13.42
Fifth grade	36	11.58	14	9.40
Sixth grade	11	3.54	6	4.03
Special education disability				
No	243	78.14		
Yes	54	17.36		
Missing	14	4.50		
Ethnicity				
White/Caucasian	133	42.77	124	83.22
Black/African American	124	39.87	16	10.74
Hispanic/Latino	43	13.83	4	2.68
Asian/Pacific Islander	1	0.32	2	1.34
Other	4	1.29	3	2.01
Missing	6	1.93	0	0.00
Gender				
Male	225	72.35	7	4.70
Female	86	27.65	142	95.30
Experimental condition				
Intervention	170	54.66	78	52.35
Comparison	141	45.34	71	47.65
Education level				
Bachelor's degree			65	43.62
Master's degree			69	46.31
Other			6	4.70
Missing			8	5.37
Observed subject				
Language Arts			86	57.72
Math			55	36.91
Social Studies			3	2.01
Science			2	1.34
Other			3	2.01
Teaching experience (years)			M = 9.20 (SD = 9.04)	

Procedures

Researchers across the three sites met with respective local school districts to recruit schools to participate in the RCT. As schools were recommended as a good fit for the study, researchers presented an overview of the RCT to the school faculty and asked for teacher volunteers. After completing informed consent procedures as outlined by university and school district institutional review boards (IRBs), volunteer teachers completed a multi-step process for study inclusion. First, teachers identified the teaching subject they associated with the worst behavior problems and used the Systematic Screening for Behavioral Disorders (SSBD; Walker & Severson, 1992) Stage 1 to nominate students whom they considered to be at risk for EBD. These nominated students and their parents completed informed assent/consent procedures as outlined by the respective IRBs. Teachers then completed the Social Skills Improvement System (SSIS; Gresham & Elliott, 2008) on consented students to verify their at-risk status according to national norms. Finally, to reduce rating scale bias, researchers confirmed the at-risk status of the students by direct observation using the Multi-Option Observational System for Experimental Studies (MOOSES; Tapp et al., 1995).

In the Fall of each year, researchers conducted five direct observations, on five different days, on each at-risk student over a period of 2 to 3 weeks before the intervention took place. Observations were typically conducted two times per week. However, in Year 1 of the study, only three to four baseline data points were collected on some students. An additional 10 observations took place during the intervention period (4–6 months) when approximately half of the students received the intervention and half served as a control. These intervention data are not reported in this article for considerations of clarity and readability. Data from the five baseline observations were analyzed for this study.

Measures

The SSBD is a gold-standard nationally normed screening measure for identifying students at risk for EBD. Stage 1 is a nomination process in which teachers study the definitions and examples of externalizing and internalizing behaviors, then consider all the students in their classroom and identify those who exhibit either or both behaviors. Once students are identified, the teacher rank orders the identified students according to the severity of behaviors. Stage 1 interrater agreement (Spearman's rho) ranges between .82 and .94, and test–retest reliability ranges between .72 and .79 (Walker & Severson, 1992).

The SSIS is a norm-referenced standardized social skills measure that includes social skills, problem behaviors, and academic scales. While teachers completed the entire 76-item measure, researchers looked at the Problem Behavior scale to verify students' at-risk status according to national norms; students whose scores fell in the at-risk or higher ranges were considered as at risk. Internal consistency (alphas) on the SSIS scales ranges from .94 to .97 (Gresham & Elliott, 2008).

MOOSES is a handheld direct observation computer system that collects data for later analysis. MOOSES has been successfully used in various studies to collect data on individual students and teachers (Downs et al., 2019; Reinke et al., 2013; Wills, Wehby, et al., 2018). With MOOSES, researchers were able to record students' disruptive behaviors and teachers' reprimands using a frequency count while simultaneously recording students' engagement as a duration event. Behavioral definitions are provided in the "Variables" section. Behaviors were coded as an occurrence any time they happened during the continual 15-min observation. Observations of a specific student occurred at approximately the same time of day during the same subject (although it varied from teacher to teacher) for all five baseline data points. Students whose MOOSES engagement levels were below 75% or whose disruptive behaviors were above 10 occurrences for a minimum of two baseline observation sessions were considered at risk, as similar characteristics have been found in other studies of behaviorally at-risk students (see, for example, Kamps et al., 2015; Wills, Kamps, et al., 2018).

Observer preparation for using the direct observation system included several steps: (a) studying the MOOSES definitions for the frequency and duration event codes, (b) coding videos of student behavior (previously scored by researchers) to 85% accuracy, and (c) observing student behavior in live non-study classrooms with a research coordinator until 85% accuracy was achieved. Observers included undergraduate and graduate students as well as researchers and research coordinators with master's or doctorate degrees.

Variables

This study focused on three target variables—teacher reprimands, student disruptive behavior, and student engagement—and their corresponding effects, after controlling for teacher praise. These variables were collected in 15-min sessions using the MOOSES program.

Teacher reprimands were defined as negative verbal comments directed toward the target student, or the group that included the target student. These comments included scolding, disapproving of student social or academic behavior, or using negative comments directing students to stop misbehavior, as well as redirection or warning of negative consequences by the teacher. For example, "Johnny, quit wasting time and get back to work," "Group three, start paying attention or your names are going on the board," and "Stop bothering Kim" would count as reprimands while "Try harder on your math worksheet; I know you can do better," "This is incorrect," and students being asked to "sit" when coming in from lunch are not recorded as reprimands.

Student disruptive behaviors were defined as deliberate physical or verbal displays of inappropriate behavior. These included posturing, gestures, and verbal statements intended to provoke others or draw attention to self; inappropriate use of classroom materials; or behavior intended to annoy, pester, or make fun of another. Disruptive behavior was any action by the student that interfered with his or her classroom participation and/or the productive classroom activity of peers. For example, rocking in the desk chair, tapping pencils, tossing material at other students or around the classroom, popping gum loudly, talking out after specifically being forbidden by teacher instructions, verbally bothering or making fun of someone, threatening ("I'm going to cut you!"), protesting ("Hey, that's not fair!"), or refusing teacher direction ("No, I won't do it," or "Make me!") would be considered disruptive behavior, while kneeling on a chair to reach the table or desk that is difficult to reach when sitting, asking a peer for a pencil or something related to assignment and then getting started on work right away, or whispering to self about instruction would not be coded as disruptive behavior.

Student engagement behaviors were defined as a student appropriately working on the assigned/approved activity. Examples of student engagement included reading orally or silently as directed, answering questions, writing, and looking at or attending to the learning material and/or task. For example, when the student is reading out loud with the class when directed to do so, or quietly following along in the

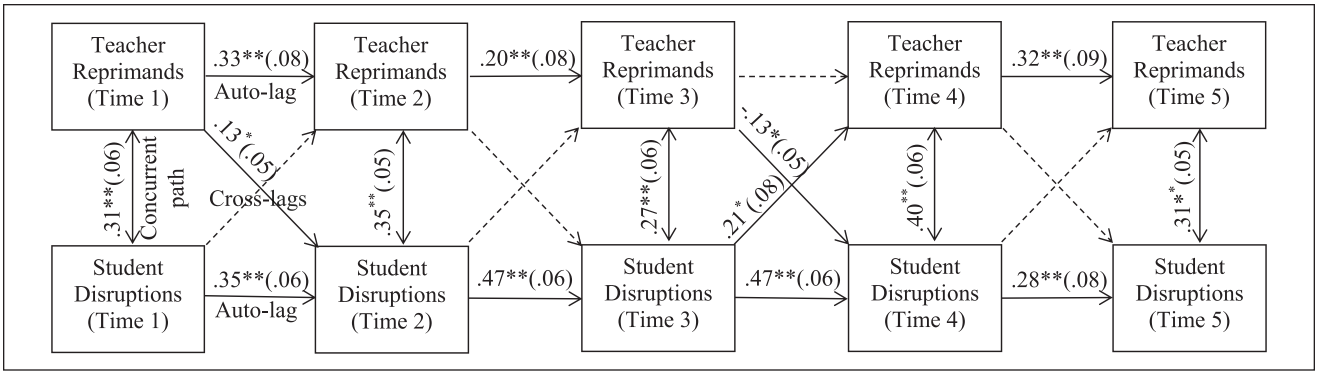

Figure 1. Teacher reprimands and student disruptions cross-lag model. Concurrent paths are correlation coefficients estimated by the model. The auto-lags and cross-lags are standardized betas. (.xx) refers to the standard error. Model controlled for grade level, site, student ethnicity, teacher ethnicity, teacher praise, the nesting effect of teachers, and time of observation. Dashed paths are not significant. To improve readability, student engagement and teacher praise auto-lags and cross-lags (while included in the model) are not shown.
$^*p < .05$. $^{**}p < .01$.

book, they would be engaged. If the student stares away from the teacher, a student responding, or instructional materials for more than 5 s, they would be disengaged.

The control variable of teacher praise was defined as a verbal statement that indicated approval of behavior over and above the evaluation of adequacy or acknowledgment of a correct response. Praise included requests for children to give themselves a pat, high five, and so on. To be counted as a praise to the student, it must be a verbal praise statement to that student or to a small or large group inclusive of the student. Examples include "Billy, I like the way you did that sum!"; "Everyone is sitting quietly, great!"; and "Team 3 is doing a great job of following directions and reading their books as I asked; excellent job!" Non-examples include the teacher saying "Thank you" to the student as she collects an assignment, "I've got Johnny's paper," and giving a thumbs up to the student.

Interobserver Agreement

During the RCT, the second trained observer was present for approximately 26.9% of observation sessions. Interobserver agreement (IOA) was collected by methods outlined by MacLean et al. (1985) using a 5-s window around each frequency code found in a primary observer's MOOSES code file. An agreement was scored on that code if the matching code was found in the second observer's file within the same 5-s window. Percentages of total agreement were calculated for frequency counts of teacher reprimands and student disruptions using the formula (agreements ÷ [agreements + disagreements]) × 100. The duration of student engagement behavior was calculated with second-by-second reliability estimates using the same IOA formula. During the RCT, overall average IOA for teacher reprimands totaled 97% (SD = 0.12), student disruptive behaviors agreed at 95% (SD = 0.11), and student engagement was 98% (SD = 2.90).

Analytical Strategy

One cross-lag model was run to isolate the causal effects of the variables of interest (teacher praise [as a control], teacher reprimands, student disruptions, and student engagement) as displayed in Figure 1. Cross-lag models have long been considered a better alternative than multiple regression in isolating causal relationships (Kenny, 1975). As the data are longitudinal, collected over 2 to 3 weeks, the theory states that we are controlling for any unobserved variables through the previous observations and the auto-lag. The cross-lags thus should be considered causal and understood as such (Kenny, 1975). Autoregressive cross-lagged models examining the association between the variables of interest (teacher praise [as a control], teacher reprimands, student disruptions, and student engagement) across five time points (at varying intervals) were estimated. All variables have cross-lags with all other variables in the model. By modeling the sequential reciprocal relationship of variables across multiple observation periods, this approach clarifies whether one variable is a leading or lagging indicator of the others. The model simultaneously estimates three different types of association pathways: autoregressive, cross-lagged, and concurrent.

The autoregressive pathways estimated the association of teacher praise at time t and at time $t + 1$ (as a control), teacher reprimands at time t and at time $t + 1$, student disruptions at time t and at time $t + 1$, and student engagement at time t and at time $t + 1$. The autoregressive pathways control for previous levels of the variable such that cross-lagged associations in our model focus on predicting change in the variable above previous levels. The cross-lagged

Table 2. Descriptive Statistics for Variables of Interest.

	Time 1		Time 2		Time 3		Time 4		Time 5	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Teacher reprimands	2.43	2.56	2.38	2.56	2.20	2.27	2.46	2.67	2.37	2.67
Student disruptions	14.32	12.67	14.65	11.70	15.90	14.11	15.71	12.98	16.46	14.33
Student engagement	70.40	22.74	67.94	22.84	67.68	23.18	67.17	23.91	65.06	24.12
Teacher praise	1.15	1.71	0.89	1.36	1.06	2.04	0.98	1.49	0.93	1.45
n	311		311		309		265		222	
% missing	0.00		0.00		0.60		14.79		28.62	

Note. Teacher reprimands, student disruptions, and teacher praise are frequency counts per 15-min interval. Student engagement is a duration event during a 15-min interval and is reported as a percentage.

pathways represent associations between teacher reprimands at time t and student disruptions at time $t + 1$; associations between student disruptions at time t and teacher reprimands at time $t + 1$; and their respective associations with student engagement, and teacher praise (as a control) across the same time points. Concurrent residual correlations between teacher reprimands, student disruptions, student engagement, and teacher praise (as a control) for each time point were also estimated. A more technical description of the model specification is included in the appendix. This modeling approach helps test aspects of the operant framework assumption that reprimands will lessen future disruptive behavior.

Time-invariant covariates were (a) site of implementation, (b) student ethnicity (White, Black, or Other), and (c) teacher ethnicity (White, Black, or Other), whereas the time-varying covariate was the time of observation. The baseline cross-lagged model was estimated with time-varying and time-invariant covariates, and all autoregressive, cross-lagged, and concurrent pathways were free to vary across time. The independence of observations assumption was violated as students were nested within classrooms; thus, a random effect of the classroom was added to the model. The school was another level of potential clustering, but as we had a relatively small number of schools ($N = 19$) from an analytic perspective, nesting at this level was not considered. The intraclass correlation coefficients (ICCs) of the variables of interest at the teacher level were moderate (approximately .35 across all time points and across all variables) and much lower at the school level (approximately .05 across all time points and across all variables) suggesting that nesting at the teacher level was necessary, whereas nesting at the school level was not. The maximum likelihood with robust standard errors (MLR) using a sandwich estimator to correct for any non-normality in the data was used to estimate the parameters and standard errors (Yuan & Bentler, 2000). The large sample size also helps resolve any normality issues. In addition, participants were students at risk, which explains why a large spike of observations at zero was not seen. Missing data were handled via the full information maximum likelihood (FIML) because it is more effective than traditional methods (listwise deletion, mean imputation) in handling missing data (Enders & Bandalos, 2001). All analyses were done in Mplus 8.4 (Muthén & Muthén, 1998–2017).

Results

Descriptive statistics for the variable of interest (teacher reprimands, student disruptions, student engagement, and the control variable of teacher praise) are shown in Table 2. There were no missing data on these variables of interest at Times 1 and 2, minimal missing at Time 3 (0.60%), 14.79% missing at Time 4, and 28.62% missing at Time 5. Bivariate correlations of the cross-section of Time 1 between the variables of interest are found in Table 3. Teacher reprimands were positively correlated with student disruptions ($r = .34$, $p < .01$) and student Black ethnicity ($r = .18$, $p < .01$). Teacher reprimands were negatively correlated with student engagement ($r = −.16$, $p < .01$), student White ethnicity ($r = −.12$, $p < .05$), grade level ($r = −.23$, $p < .01$), and the Utah site status ($r = −.19$, $p < .01$). Students' disruptions were positively correlated with teacher White ethnicity ($r = .12$, $p < .05$) and negatively correlated with student engagement ($r = −.30$, $p < .01$), grade level ($r = −.20$, $p < .01$), and the Utah site status ($r = −.12$, $p < .01$). Student engagement was positively correlated with praise ($r = .17$, $p < .01$) and the Utah site status ($r = .21$, $p < .01$) and negatively correlated with Missouri site status ($r = −.28$, $p < .01$). Praise was not positively correlated with any variables but was negatively correlated with grade level ($r = −.11$, $p < .05$).

The results of the cross-lag model showing standardized results are reported in Table 4 and illustrated in Figure 1. The auto-lags and cross-lags were freely estimated across time. Not shown are the controls of time-invariant covariates of student ethnicity, teacher ethnicity, grade level, nesting effect of teachers, site status, and the time-varying covariate of observation time.

Table 3. Time One Bivariate Correlations of Variables of Interest ($N = 311$).

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9	10.	11.	12.
1. Teacher reprimands	1.00											
2. Student disruptions	.34**	1.00										
3. Student engagement	−.16**	−.30**	1.00									
4. Teacher praise	.03	.01	.17**	1.00								
5. Student ethnicity (White)	−.12*	−.00	.02	−.01	1.00							
6. Student ethnicity (Black)	.18**	.08	−.00	−.02	−.70**	1.00						
7. Teacher ethnicity (White)	−.08	.12*	−.10	.01	.20**	−.29**	1.00					
8. Teacher ethnicity (Black)	.03	−.06	.08	−.07	−.19**	.29**	−.78**	1.00				
9. Grade level	−.23**	−.20**	.01	−.11*	.010	−.06	.03	.02	1.00			
10. Utah (site)	−.19**	−.12**	.21**	−.01	.33**	−.55**	.22**	−.23**	.20**	1.00		
11. Missouri (site)	.09	.07	−.28**	−.00	−.01	.17**	.07	−.04	.05	−.56**	1.00	
12. Tennessee (site)	.10	.05	.10	.01	−.34**	.39**	−.31	.28**	−.27**	−.44**	−.50**	1.00

*$p < .05$. **$p < .01$.

Table 4. Standardized Concurrent and Cross-Lagged Paths as Produced by the Model.

Variable	Disrupt Time 1	Disrupt Time 2	Disrupt Time 3	Disrupt Time 4	Disrupt Time 5	Engage Time 1	Engage Time 2	Engage Time 3	Engage Time 4	Engage Time 5
Reprimands Time 1	**.31****	.13*	—	—	—	**−.15****	−.01	—	—	—
Reprimands Time 2	—	**.35****	−.02	—	—	—	**−.16****	.09	—	—
Reprimands Time 3	—	—	**.27****	−.13*	—	—	—	**−.20****	.06	—
Reprimands Time 4	—	—	—	**.40****	.11	—	—	—	**−.22****	−.08
Reprimands Time 5	—	—	—	—	**.31****	—	—	—	—	**−.20****
Praise Time 1	**−.01**	.06	—	—	—	**.19****	.02	—	—	—
Praise Time 2	—	**−.00**	−.05	—	—	—	**.20****	.07	—	—
Praise Time 3	—	—	**.05**	.02	—	—	—	**.09**	−.01	—
Praise Time 4	—	—	—	**.13***	.11	—	—	—	**.11***	.08
Praise Time 5	—	—	—	—	**−.04**	—	—	—	—	**.12**

Note. All bolded diagonal values are concurrent paths or correlation coefficients estimated by the model. All other values are cross-lags or standardized betas. The model controlled for grade level, site, student ethnicity, teacher ethnicity, the nesting effect of teachers, and time of observation. "—" = not applicable; Disrupt = student disruptions; Engage = student engagement; Reprimands = teacher reprimands; Praise = teacher praise.
*$p < .05$. **$p < .01$.

Teacher Reprimands and Student Disruptions

Teacher reprimands were weakly predictive of future teacher reprimands after controlling for teacher praise, except for Time 3 predicting Time 4 (see Figure 1), which was not entirely unexpected given the pattern of missing data as explained by the differences in the data collection across the years as discussed above. Disruptions were also weakly predictive of future disruptions across all time points. The R^2 values for student disruptions were 7.60% for Time 1, 22.60% for Time 2, 26.70% for Time 3, 25.00% for Time 4, and 28.50% for Time 5. The jump in R^2 from Time 1 to the other time points shows the strength of the cross-lag modeling approach, where previous time points are accounted for allowing for stronger causal inferences to be made.

Most of the cross-lags themselves were non-significantly predictive of future time points with the exception of teacher reprimands at Time 1 positively predicting student disruptions at Time 2 ($\beta = .13$, $p < .05$), student disruptions at Time 3 positively predicting teacher reprimands at Time 4 ($\beta = .21$, $p < .05$), and teacher reprimands at Time 3 negatively predicting student disruptions at Time 4 ($\beta = -.13$, $p < .05$). These latter cross-lags have fewer causal implications because of the non-significant auto-lag of teacher reprimands between Time Points 3 and 4. The concurrent correlations between teacher reprimands and student disruptions were all positive, statistically significant, and small. Of note, the concurrent correlation at Time 1 was close but not the same as is shown in Table 3. This is to be expected as the concurrent correlations were estimated in the context of the cross-lag model.

Teacher Reprimands and Student Engagement

The last two research questions addressed whether teacher reprimands would affect student engagement (or the reciprocal) after controlling for teacher praise. In our analysis, teacher reprimands had no significant cross-lags associated with student engagement or vice versa. These results are therefore not shown.

Discussion

Reprimands, often intended as a form of punishment, are used more often than praise in schools (Reinke et al., 2013; Van Acker et al., 1996). Despite many teachers' tendencies to use them in response to student disruptive behavior (Hollingshead et al., 2016), reprimands have been linked to escape-motivated behaviors (Shores et al., 1993), aggression (Van Acker et al., 1996), and further disruptive behavior (Downs et al., 2019). The use of reprimands for students with or at risk for EBD can be especially problematic, given the specific school challenges faced by these students (Conley et al., 2014; Hecker et al., 2014; Salle et al., 2018). Unfortunately, many students with EBD receive high rates of reprimands and low rates of praise from their teachers (Rathel et al., 2014; Sutherland & Wehby, 2001), although it would be beneficial for these students to receive fewer reprimands and more praise to improve their engagement and academics (Downs et al., 2019; Rathel et al., 2014).

The current study found that teacher reprimands did not appear to decrease future disruptive behavior or increase future engagement for students at risk for EBD or vice versa. Our results support earlier assertions that reprimands do not result in long-term positive behavior change, although they may temporarily suppress misbehavior (Alber & Heward, 2000). This might be because reprimands do not directly teach students the skills needed to improve their behavior (Curran & IRIS Center, 2003; Taylor et al., 2009), and thus students may continue to exhibit negative behavior and continue receiving reprimands. Another problem is that reprimands are reactive: a student acts disruptively and a teacher reprimands the student. We agree with others who recommend effective teaching techniques and proactive behavior management strategies (Simonsen et al., 2008; Taylor et al., 2009) to decrease disruptions and increase engagement. It was also interesting that student misbehavior did not predict future teacher reprimands. This may be explained by the possibility that teachers are not reinforced for their reprimands, in that, as shown in the present study, reprimanding students has no lasting effects on student misbehavior. Also, while teacher reprimands were negatively correlated with student engagement, they did not predict future student engagement for these students at risk for EBD. A possible reason for this finding may be that teachers are more focused on reprimanding students who are disruptive rather on students who are disengaged, especially if such students are not disruptive to the classroom learning environment. Teachers may use different behavior management strategies for disengaged students such as reminding students of expectations or reinforcing peers who are engaged, although this finding needs additional study.

The small correlations between teacher reprimands and student disruptions found in this study were somewhat unexpected. These correlations were expected to be larger, although there are many reasons why they may have been smaller. For example, the teacher may not catch all the disruptive behavior that an observer may see, the teacher may mistakenly reprimand the student, or the student may be disengaged but not be disrupting the class. It is also possible that a student could receive a group reprimand, because they were part of a group that was being disruptive, although the student themselves was not disruptive.

While not a focus of the study, our results suggest that as student grade level increases, teachers' verbal interactions with students (i.e., reprimands and praise) tend to decrease, similar to results found by Reddy et al. (2013) in their study of elementary school classrooms. These results are concerning since research suggests that students with or at risk for EBD benefit from high rates of teacher praise and low rates of teacher reprimands (Downs et al., 2019; Rathel et al., 2014; Reinke et al., 2013).

There were some other interesting findings from this study. Namely, the Utah site had significantly fewer teacher reprimands, student disruptions, and greater student engagement than the other sites. The Missouri site had significantly less student engagement than the other sites. Reasons for these site differences are unclear but may be due to cultural or demographic differences—further research would be needed to investigate these findings. In addition, there were patterns found regarding student and teacher minority status. White students were reprimanded significantly less ($r = -.12$) than Black students or other students. Black students were reprimanded significantly more ($r = .18$) than White students and other students. For teachers, there were more student disruptions if the teacher was White ($r = .12$) than if the teacher was Black or Other. These results are like those of Scott et al. (2019) who found that both Black and White teachers provided significantly more negative feedback to Black students (than White students) regardless of their behavior, suggesting a need for interventions to address these differences.

Limitations and Future Directions

Although participant data were from an RCT, the sample size was relatively limited by observing and recording only students at risk for EBD, not all the students in the class. We analyzed the data from students at risk for EBD, as we hypothesized they would receive more reprimands than their peers, as indicated in the literature (Rathel et al., 2014; Sutherland & Wehby, 2001). Future studies would benefit from examining the teacher reprimands, student disruptions, and student engagement levels for all students in a class.

The models used in the current study did not control for other teacher behaviors that may confound future student disruptive behaviors. Student behavior may be maintained by other variables such as non-verbal reprimands, student–teacher relationships, or token economies. Several studies could be conducted to separately examine these variables. For example, observations could be conducted that examine non-verbal reprimands (disapproving looks, tickets, name written on board, hand signals, etc.), the effects of negative consequences ("clip down," time out, lost recess time, principal's office), and other variables (student–teacher relationship, teacher job satisfaction or burnout) on students' disruptive classroom behavior.

In addition, the small predictive auto-lags, the power of the model, turned out to be a limitation, as we would expect the auto-lags to be more predictive from one time point to the next. However, this may be influenced by the varying numbers of days between observations, as data were not collected daily, neither were observations consistently spaced across the 2- to 3-week observational period. Another possibility for the weak auto-lags (especially between Time Points 3 and 4) is the increased missingness in the variables of interest that occurred as a feature of this study as discussed in the "Method" section. Future studies might find that data gathered daily will reveal stronger relationships between the auto-lags, as well as verifying whether the cross-lags are more predictive. Finally, data were only collected in elementary school settings. Replication of this study in secondary school settings would be beneficial.

Implications and Conclusions

Reprimands are meant to stop misbehavior. In the current study, teacher reprimands did not appear to help decrease future classroom disruptions or increase future engagement of students at risk for EBD. This should not be surprising, as harsh reprimands in schools have been associated with negative side effects such as anger, fear, escape, and avoidance (Sidman, 1989; Walker et al., 2004) rather than improved student behavior. In addition to being harmful to teachers and their students, reprimands prove less effective than positive classroom behavioral management strategies (Reinke et al., 2013). Teachers who use reprimands also report higher levels of emotional exhaustion than their peers who do not (Reinke et al., 2013). Given the findings of the current study, along with those of previous researchers, we recommend that teachers replace reprimands with proactive classroom management strategies, such as clearly teaching classroom expectations, reinforcing positive student behavior, and using behavior-specific praise (Simonsen et al., 2008), as primary responses to student misbehavior and disengagement.

Appendix

This article uses an autoregressive cross-lag model to understand the relationships of teacher reprimands, student disruptive behavior, and student engagement. This is a very brief explanation of the purpose and rationale for using the cross-lag model. Readers interested in an in-depth exploration of this model are encouraged to read the work of Kenny (1975). The general model has the following form where Y_{ti} and X_{ti} are ith student disruptions and teacher reprimands at time t, respectively:

$$
\begin{aligned}
Y_{ti} &= \beta_{0y} + \lambda_y Y_{(t-1)i} + \beta_x X_{(t-1)i} + \delta_{ti} \\
X_{ti} &= \beta_{0x} + \lambda_x X_{(t-1)i} + \beta_y Y_{(t-1)i} + \varepsilon_{ti},
\end{aligned}
$$

where $Y_{(t-1)i}$ and $X_{(t-1)i}$ are those same scores but recorded at time $t-1$. The λ_y and λ_x terms are the autoregressive lags of the scores at the previous time point. As these terms are included in the model, greater confidence in the lack of bias of the cross-lag parameters (β_x and β_y) can be given. These cross-lag parameters are estimates of which variable causes change in the other. If β_x is significant, then there is evidence that teacher reprimands lead to student disruptions, as it determines part of the variance of future student disruptions. The δ_{ti} and ε_{ti} terms are the error terms for student disruptions and teacher reprimands at time t.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported in this article was supported in part by a grant from the Institute of Education Sciences and the U.S. Department of Education (Grant No. R324A120344), awarded to the University of Kansas in cooperation with Brigham Young University and Vanderbilt University. The opinions presented in this article are those of the authors, and no endorsement by the agency is intended or implied.

ORCID iD

Paul Caldarella https://orcid.org/0000-0002-0883-8890

References

Alber, S., & Heward, W. (2000). Teaching students to recruit positive attention: A review and recommendations. *Journal of Behavioral Education*, *10*, 177–204. https://doi.org/10.1023/a:1012258231074

Allday, R. A., Hinkson-Lee, K., Hudson, T., Neilsen-Gatti, S., Kleinke, A., & Russel, C. S. (2012). Training general educators to increase behavior-specific praise: Effects on students with EBD. *Behavioral Disorders*, *37*, 87–98. https://doi.org/10.1177/019874291203700203

Allday, R. A., & Pakurar, K. (2007). Effects of teacher greetings on student on-task behavior. *Journal of Applied Behavior Analysis*, *40*, 317–320. https://doi.org/10.1901/jaba.2007.86-06

Caldarella, P., Larsen, R. A. A., Williams, L., Downs, K. R., Wills, H. P., & Wehby, J. H. (2020). Effects of teachers' praise-to-reprimand ratios on elementary students' on-task behaviour. *Educational Psychology*. Advance Online Publication. https://doi.org/10.1080/01443410.2020.1711872

Caldarella, P., Larsen, R. A. A., Williams, L., Wills, H. P., & Wehby, J. H. (2019). Teacher praise-to-reprimand ratios: Behavioral response of students at risk for EBD compared with typically developing peers. *Education and Treatment of Children*, *42*, 447–468. https://doi.org/10.1353/etc.2019.0021

Chafouleas, S. M., Volpe, R. J., Gresham, F. M., & Cook, C. R. (2010). School-based behavioral assessment within problem-solving models: Current status and future directions. *School Psychology Review*, *39*, 343–349.

Conley, L., Marchant, M., & Caldarella, P. (2014). A comparison of teacher perceptions and research-based categories of student behavior difficulties. *Education*, *134*, 439–451.

Cook, C. R., Fiat, A., Larson, M., Daikos, C., Slemrod, T., Holland, E. A., . . . Renshaw, T. (2018). Positive greetings at the door: Evaluation of a low-cost, high-yield proactive classroom management strategy. *Journal of Positive Behavior Interventions*, *20*, 149–159. https://doi.org/10.1177/1098300717753831

Curran, C., & The IRIS Center. (2003). *Encouraging appropriate behavior*. IRIS, Peabody, Vanderbilt University. http://iris.peabody.vanderbilt.edu/wp-content/uploads/pdf_case_studies/ics_encappbeh.pdf

Downs, K. R., Caldarella, P., Larsen, R. A. A., Charlton, C. T., Wills, H. P., Kamps, D. M., & Wehby, J. H. (2019). Teacher praise and reprimands: The differential response of students at risk for emotional and behavioral disorders. *Journal of Positive Behavior Interventions*, *21*, 135–147. https://doi.org/10.1177/1098300718800824

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*, 430–457. https://doi.org/10.1207/s15328007sem0803_5

Gable, R. A., Hester, P. P., Rock, M. L., & Hughes, K. (2009). Back to basics: Rules, praise, ignoring, and reprimands revisited. *Intervention in School and Clinic*, *44*, 195–205. https://doi.org/10.1177/1053451208328831

Gresham, F. M., & Elliott, S. N. (2008). *Social Skills Improvement System: Rating Scales*. Pearson Assessments.

Hecker, B. L., Young, E. L., & Caldarella, P. (2014). Teacher perspectives on behaviors of middle and junior high school students at risk for emotional and behavioral disorders. *American Secondary Education*, *42*(2), 20–32.

Hollingshead, A., Kroeger, S. D., Altus, J., & Trytten, J. B. (2016). A case study of positive behavior supports-based interventions in a seventh-grade urban classroom. *Preventing School Failure: Alternative Education for Children and Youth*, *60*(4), 1–8. https://doi.org/10.1080/1045988X.2015.1124832

Jennings, P. A., & Greenberg, M. T. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research*, *79*, 491–525. https://doi.org/10.3102/0034654308325693

Kamps, D., Conklin, C., & Wills, H. (2015). Use of self-management with the CW-FIT group contingency program. *Education and Treatment of Children*, *38*, 1–32. https://doi.org/10.1353/etc.2015.0003

Kenny, D. A. (1975). Cross-lagged panel correlation: A test for spuriousness. *Psychological Bulletin*, *82*, 887–903. https://doi.org/10.1037/0033-2909.82.6.887

Kodak, T., Northup, J., & Kelley, M. E. (2007). An evaluation of the types of attention that maintain problem behavior. *Journal of Applied Behavior Analysis*, *40*, 167–171. https://doi.org/10.1901/jaba.2007.43-06

MacLean, W. E., Tapp, J. T., & Johnson, W. L. (1985). Alternate methods and software for calculating interobserver agreement for continuous observation data. *Journal of Psychopathology and Behavioral Assessment*, *7*, 65–73. https://doi.org/10.1007/bf00961847

Martens, B. K., & Ardoin, S. P. (2010). Assessing disruptive behavior within a problem-solving model. In G. G. Peacock, R. A. Ervin, E. J. Daly & K. W. Merrell (Eds.), *Practical handbook of school psychology: Effective practices for the 21st century* (pp. 157–174). Guilford Press.

McComas, J. J., Downwind, I., Klingbeil, D. A., Petersen-Brown, S., Davidson, K. M., & Parker, D. C. (2017). Relations between instructional practices and on-task behavior in classrooms serving American Indian students. *Journal of Applied School Psychology*, *33*, 89–108. https://doi.org/10.1080/15377903.2016.1236308

Merrell, K. W., Ervin, R. A., & Gimpel Peacock, G. (2012). *School psychology for the 21st century: Foundations and practices* (2nd ed.). The Guilford Press.

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.).

Rathel, J. M., Drasgow, E., Brown, W. H., & Marshall, K. J. (2014). Increasing induction-level teachers' positive-to-negative communication ratio and use of behavior-specific praise through e-mailed performance feedback and its effect on students' task engagement. *Journal of Positive Behavior Interventions*, *16*, 219–233. https://doi.org/10.1177/1098300713492856

Rathel, J. M., Drasgow, E., & Christle, C. C. (2008). Effects of supervisor performance feedback on increasing preservice teachers' positive communication behaviors with students with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders*, *16*, 67–77. https://doi.org/10.1177/1063426607312537

Reddy, L. A., Fabiano, G. A., Dudek, C. M., & Hsu, L. (2013). Instructional and behavior management practices implemented by elementary general education teachers. *Journal of School Psychology*, *51*, 683–700. https://doi.org/10.1016/j.jsp.2013.10.001

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*, 39–50. https://doi.org/10.1177/1098300712459079

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. https://doi.org/10.1037/a0022714

Salle, T. L., George, H. P., McCoach, D. B., Polk, T., & Evanovich, L. L. (2018). An examination of school climate, victimization, and mental health problems among middle school students self-identifying with emotional and behavioral disorders. *Behavioral Disorders*, *43*, 383–392. https://doi.org/10.1177/0198742918768045

Scott, T. M., Gage, N., Hirn, R., & Han, H. (2019). Teacher and student race as a predictor for negative feedback during instruction. *School Psychology*, *34*, 22–31. https://doi.org/10.1037/spq0000251

Sherrill, J. T., O'Leary, S. G., Albertson-Kelly, J. A., & Kendziora, K. T. (1996). When reprimand consistency may and may not matter. *Behavior Modification*, *20*, 226–236. https://doi.org/10.1177/01454455960202006

Shores, R. E., Gunter, P. L., & Jack, S. L. (1993). Classroom management strategies: Are they setting events for coercion? *Behavioral Disorders*, *18*, 92–102. https://doi.org/10.1177/019874299301800207

Sidman, M. (1989). *Coercion and its fallout*. Authors Cooperative.

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

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. Appleton-Century.

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. https://doi.org/10.1177/002246690103500306

Tapp, J., Wehby, J., & Ellis, D. (1995). A multiple option observation system for experimental studies: MOOSES. *Behavior Research Methods, Instruments, & Computers*, *27*, 25–31. https://doi.org/10.3758/BF03203616

Taylor, R. L., Smiley, L. R., & Richards, S. B. (2009). *Exceptional students: Preparing teachers for the 21st century*. McGraw-Hill Higher Education.

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

Walker, H. M., Ramsey, E., & Gresham, F. (2004). *Antisocial behavior in school* (2nd ed.). Wadsworth Thomson Learning.

Walker, H. M., & Severson, H. H. (1992). *Systematic Screening for Behavior Disorders* (2nd ed.). Sopris West.

White, M. A. (1975). Natural rates of teacher approval and disapproval in the classroom. *Journal of Applied Behavior Analysis*, *8*, 367–372. https://doi.org/10.1901/jaba.1975.8-367

Wills, H., Wehby, J., Caldarella, P., Kamps, D., & Swinburne Romine, R. (2018). Classroom management that works: A replication trial of the CW-FIT program. *Exceptional Children*, *84*, 437–456. https://doi.org/10.1177/0014402918771321

Wills, H. P., Kamps, D., Caldarella, P., Wehby, J., & Swinburne-Romine, R. S. (2018). Class-Wide Function-Related Intervention Teams "CW-FIT": Outcomes from a multi-site randomized replication trial. *The Elementary Education Journal*, *119*, 29–51. https://doi.org/10.1086/698818

Wills, H. P., Kamps, D., Hansen, B. D., Conklin, C., Bellinger, S., Neaderhiser, J., & Nsubuga, B. (2010). The classwide function-based intervention team program. *Preventing School Failure: Alternative Education for Children and Youth*, *54*, 164–171. https://doi.org/10.1080/10459880903496230

Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. In M. E. Sobel & M. P. Becker (Eds.), *Sociological methodology 2000* (pp. 165–200). ASA.