How Does Collective Student Behavior and Other Classroom Contextual Factors Relate to

Teachers' Implementation of an Evidence-Based Intervention? A Multilevel SEM

Rashelle J. Musci[1], Elise T. Pas[1], Amie F. Bettencourt[1,2], Katherine E. Masyn[3],

Nicholas S. Ialongo[1], & Catherine P. Bradshaw[1,4]

[1]Johns Hopkins University, Bloomberg School of Public Health

[2] Johns Hopkins University, School of Medicine

[3]Georgia State University, School of Public Health

[4]Unviersity of Virginia, Curry School of Education and Human Development

Published in *Development and Psychopathology*

Abstract

Building on prior work regarding the potential for peer contagion or deviance training in group delivered interventions (Dishion & Dodge, 2005, 2006; Dodge, Dishion, & Lansford, 2006), we leveraged data from a randomized trial, testing the integration of two preventive interventions (PATHS and PAX Good Behavior Game; GBG), to explore the extent to which classroom contextual factors served as either a barrier to or a motivator for teachers to implement the evidence-based PAX GBG with high frequency or dosage. We included students' baseline levels of behavior, measured both with regards to positive (i.e., engagement, social emotional skills) and negative (i.e., hyperactive and aggressive-disruptive) behaviors. Data were collected from 204 teachers in 18 urban elementary schools. A series of multilevel structural equation models (ML-SEM) were fit to the data. The analyses indicated that classrooms with higher classroom levels of aggressive behavior, on average, at baseline had teachers with lower implementation dosage (i.e., played fewer games) across the school year. In addition, teachers who reported higher baseline levels of emotional exhaustion, regardless of student behavior, also reported lower implementation dosage. Taken together, the results indicated that negative, but not positive, contextual factors at baseline were related to lower implementation dosage; this, in turn, suggests that negative contextual factors may serve as a barrier, rather than a motivator, of teachers' implementation dosage of classroom-based preventive interventions.

How Does Collective Student Behavior and Other Classroom Contextual Factors Relate to Teachers' Implementation of an Evidence-Based Intervention? A Multilevel SEM

There is a large and growing body of research focused on the potential for group-delivered interventions for aggressive/disruptive children to result in a peer contagion process, which includes deviance training, and thus can result in iatrogenic program effects (e.g., Dishion & Tipsord, 2011; Dishion & Dodge, 2005, 2006; Dodge, Dishion, & Lansford, 2006). Although much of the work to date has focused on group interventions and group dynamics, this peer contagion work has direct applicability to universal behavioral programming in classrooms. For example, it is likely that teachers' ability to deliver universal program content is especially challenging in classrooms with large numbers of students lacking inhibitory controls. In high-risk school settings, there may be a higher proportion of students with challenging behaviors, whereby the group context could similarly allow for the collective modeling and reinforcement of deviant student behaviors. Also applicable in classrooms is the coercion theory of antisocial behavior, wherein parent inability to effectively manage child non-compliance can trigger a coercive cycle of escalating aversive behavior on the part of the both the parent and the child (Patterson, Reid, & Dishion, 1992). A similar coercive cycle may develop in classrooms as well, even when teachers attempt to implement evidence-based classroom behavior management programs.

In the current study, we extended Dishion and colleagues expansive line of work on peer deviance training to the elementary school classroom context by examining the association between classroom-level aggressive-disruptive behavior and teachers' implementation of the PAX Good Behavior Game (PAX GBG; Embry et al. 2003; Ialongo et al., 2019), which is a classroom-wide intervention designed to reduce aggressive/disruptive behavior. We also

examined whether positive and prosocial student behavior was associated with implementation. Although there has been considerable research documenting the potential impact of classroom-based preventive interventions (e.g., Chiapa et al., 2016; Smith et al., 2018), there remains a gap in the consideration of classroom behavioral dynamics in relation to implementation, which in turn might contribute to the overall impact of these programs. This study aims to fill this gap, with the overarching goal of determining the context and conditions under which a commonly used classroom-based prevention intervention (i.e., PAX GBG) is most effective at addressing aggressive/disruptive behavior problems.

Association between Context and Implementation Fidelity

The field of implementation science continues to broaden, encompassing a plethora of conceptual theories, frameworks, and models (Nilsen, 2015) defining a myriad of implementation-relevant factors (e.g., Durlak & DuPre, 2008; Han & Weiss, 2005) at multiple contextual levels as well as the corresponding supports needed (e.g., Domitrovich et al., 2008). Also recognized is the integral importance of practitioners and implementers to achieve implementation fidelity (e.g., Fixsen, Blase, Naoom, & Wallace, 2009; Lochman et al., 2017). Implementation studies have addressed the classroom and school contexts, aligning to the above-noted models and frameworks. However, the emphasis has been on measuring the observed or teacher-reported perceptions of school and classroom climate and teacher demographics, training and education, and other teacher factors such as efficacy and burnout. The research findings regarding the potential influence of these variables are mixed, whereby some are null, positive, or negative.

For example, prior studies of preventive interventions have suggested that factors such as administrator support (Kam, Greenberg, & Walls, 2003; Rohrbach, Graham, & Hansen, 1993)

and efficacy (e.g., Ringwalt et al., 2003; Rohrbach et al., 1993) are important predictors of

implementation. Similarly, in one study, teacher burnout was linked with lower implementation

dosage (Domitrovich et al., 2015); however, other studies have suggested that burnout may

moderate the association between coach-teacher alliance and adherence to an intervention

protocol (Wehby, Maggin, Moore Partin, & Robertson, 2012). The teachers' ratings of the social

validity of the intervention (as measured by the perception of program fit with the teacher's style

[Domitrovich et al., 2015], and as fit, effectiveness, and reduced burden [Wehby et al., 2012])

are also associated with higher implementation dosage and adherence. The findings regarding

teacher training, education, and demographics are somewhat inconclusive, whereby some studies

report null findings for education level and experience (e.g., Domitrovich, Gest, Gill, Jones, &

DeRouise, 2015; Ringwalt et al., 2003; Wanless et al., 2015), and others report a positive

association (e.g., Domitrovich et al., 2009; Sutherland et al., 2018a).

Another possible contextual influence on implementation is the classroom climate. Much

of the prior research has shown a positive association between ratings of positive school climate

and implementation (Bradshaw, Koth, Thornton, & Leaf, 2009; Malloy et al., 2015; Pas,

Waasdorp, & Bradshaw, 2015); however, some studies suggest a negative relationship over time

(Pas et al., 2015; Sutherland et al., 2018a). For example, a study testing a targeted intervention

indicated that observed emotional supports within the classroom were positively associated with

implementation competence and adherence over time, while observed classroom organization

was negatively associated with growth in adherence over time (Sutherland et al., 2018a). The

classroom organization measure in the Sutherland et al. study focused on the teachers' behavioral

and instructional management strategies. They posited that in the context of such foundational

supports, there may have been less room or need for growth in adherence to this intervention.

Taken together, extant mixed findings suggest a need for additional exploration of both teacher and classroom contextual influences on implementation (Domitrovich et al., 2009).

Student Behavior as a Contextual Factor

Few studies have considered the extent to which students' behavior in the classroom prior to program implementation may relate to teacher implementation fidelity. Disorder or dysfunction measured more broadly within other, non-classroom contexts has been shown to be associated with fidelity. A national survey of school-based interventions found that large urban schools with large proportions of poor students reported implementing more programs (Payne, Gottfredson, D. C., & Gottfredson, G. D., 2006). Further, in studies of family/home interventions, analogous variables such as familial stress levels and current mental health problems have been found to be related to higher family engagement and thus dosage received in one widely-used home visiting program, but not another (e.g., Latimore et al., 2017). These findings suggest that implementers serving environments in greater need of prevention programs may be more motivated to implement programs and, in some cases, with higher fidelity. On the other hand, research regarding the implementation of positive behavior supports has demonstrated that schools with a higher suspension rate and poorer organizational health have lower implementation (Bradshaw, Mitchell, & Leaf, 2009; Pas et al., 2015). Classrooms with greater problem behaviors may similarly suffer from poorer implementation.

Another Sutherland et al. (2018b) study recently examined ratings of student "problem" and "externalizing" behavior in relation to adherence and quality for the implementation of BEST in CLASS; their analyses suggested a positive and significant association between only problem behaviors and implementation quality. It is important to explore, however, whether these findings generalize to universal prevention programs, as compared to the targeted

intervention tested by Sutherland and colleagues (2018a and b). Exploration of a broader set of both positive and negative student behaviors at baseline may also better inform our understanding of the extent to which student behavior predicts implementation dosage of universal programs. Positive student behavior, as compared to problematic student behavior, has been largely overlooked in the literature on possible classroom level predictors of implementation.

Overview of the Current Study

In present study, we simultaneously explored two research questions regarding the direction of the potential effects of collective classroom-level student behaviors and teachers beliefs and perceptions on teacher implementation of the PAX GBG. The first research question centered on the effects of collective classroom-level behaviors, both positive and negative, on implementation, while the second was focused on the role of teacher beliefs and perceptions. With regard to aggressive/disruptive behavior, consistent with Patterson, Reid, and Dishion's (1992) coercion theory of antisocial behavior, it was expected that higher levels of classroom aggressive/disruptive behavior would be associated with lower levels of teacher implementation of the PAX GBG. A second and competing hypothesis was that higher classroom levels of aggressive/disruptive behavior may motivate teachers to implement the PAX GBG more often in order to gain better control of student behavior. Further, we reasoned that teachers in classrooms with a high baseline level of student prosocial behaviors may not feel the urgency to implement the PAX GBG at a high dosage, particularly if they feel their students already exhibit the competencies targeted by the program. Alternatively, teachers in classrooms with fewer behavioral issues may have more flexible class time to dedicate to implementing the PAX GBG and, thus, may do so more frequently.

We operationalized implementation dosage in the present study as the number of PAX GBG 'games' played across the school year and accounted for student gender as a correlate of individual student behavior. We also included teacher demographics and teacher perceptions and beliefs as correlates of teacher implementation at the classroom level, as prior research has suggested they may be predictive of implementation of PAX GBG (i.e., Domitrovich et al., 2015).

In order to address some of the methodological challenges associated with modeling student behaviors nested within classrooms, we used multilevel structural equation modeling (ML-SEM). This flexible modeling approach allows classroom-level variability in student behaviors to impact implementation at the teacher level. Previous research utilizing ML-SEM has demonstrated the ability to model and test contextual effects using individual-level data (Dunn, Masyn, Johnston, & Subramanian, 2015) and the ability to test differences in the factor structure of a construct at multiple levels (Huang & Cornell, 2016). This approach enabled us to examine between-classroom variation in collective student behaviors, while also adjusting for teacher-level characteristics that had been previously linked with implementation (e.g., demographics, burnout, organizational health). Specifically, modeing classroom-level random effects of the student-level behaviors act here as latent predictors of the classroom-level outcome of program implementation by the teacher. Multilevel latent variable models, for example, multilevel factor analysis (ML-FA), multilevel path analysis (ML-PA) and multilevel SEM (ML-SEM), are being used more frequently in multilevel studies as reflected in a recent review of the literature which described 27 such applications across multiple disciplines (Kim, Dedrick, Cao, & Ferron, 2016). However, *none* of these studies utilized this ML-SEM to explore the association between of student-level behaviors on classroom or teacher-level measures, like

implementation. This is a novel modeling approach and uniquely suited for examining our research questions, as it allows for appropriate disentanglement of within- and between-classroom variability in student behavior.

<div align="center">**Method**</div>

Design Overview

The data for this study were collected as part of a randomized controlled trial (RCT; Ialongo et al., 2019) of the integration of the PAX GBG with the Promoting Alternative THinking Strategies (i.e., PATHS; Greenberg, Kusché, & CPRG, 2011; Kusché, Greenberg & CPPRG, 2011) social-emotional learning curriculum. Participating schools were randomized to one of three conditions: the integrated model (9 schools), the PAX GBG only (9 schools), and a control condition (9 schools). Random assignment to one of the three intervention arms (control, PAX GBG only, or integrated PATHS to PAX GBG) occurred at the school level, whereby all classrooms within the school were assigned the same condition; therefore, the subsample of the 155 classrooms within the 18 intervention schools was included in the current study. Given our focus on implementation, the 9 control schools were excluded from the current analyses. Data regarding PAX GBG implemention were collected in each of the 18 intervention schools.

Participants

The current study sample included the 3,115 students who were in the 204 intervention K-5 classrooms across the 18 intervention schools; of the 204 teachers, and thus classrooms, about 60% were in schools randomly assigned to deliver PAX GBG, whereas about 40% were in schools randomized to deliver the integrated PAX GBG and PATHS curriculum. Students were predominantly African American (87.5%) and 49.1% were male. This is representative of the broader district population, which was a large, urban, east coast public school district with a

largely African American student population (88%, on average). The vast majority of sample teachers were female (i.e., 91.0%) and slightly less than half were 30 or younger and taught students in grades 3 through 5.

Intervention

Teachers in both intervention conditions were trained to implement the PAX GBG (i.e., either as a stand-alone intervention or as integrated with PATHS). The Good Behavior Game was originally developed by Barrish et al. (1969) and applies social learning principles to a team-based game that classroom teachers initiate as a means to promote "good behavior" and reduce aggressive, disruptive, and off-task behavior. In a more recent update and augmentation, Embry and colleagues' (2003) added verbal and visual cues to the "PAX GBG" version to promote generalization of increased attention and prosocial behaviors throughout the school day, and not just when playing the GBG. Various versions of the GBG have been tested over the past two decades and most have demonstrated positive effects on academic, behavioral, and substance use outcomes (see Ialongo et al., 1999; 2001).

Procedure

Recruitment for the intervention study occurred at the school level; all principals agreed to participate in the year-long project and allow their teachers to receive training and coaching in the interventions. Teacher participation was voluntary; teachers were recruited to participate by project staff and provided written consent to report on their beliefs and perceptions about teaching and workplace stress level. Parents of student participants provided active written consent for participation in the data collection (Ialongo et al., 2019). The current study utilizes secondary data analysis among intervention classrooms only; the original RCT, along with the current study, were approved by the relevant Institutional Review Boards.

Training for the interventions included a multi-day workshop at the start of school year (i.e. 1.5 days for the PAX GBG only condition and 3.5 days of training for the integrated condition, of which the same 1.5 days was focused on PAX GBG). Across the entire school year, teachers also received 31 weeks of weekly face-to-face coaching, which was a manualized coaching approach that included tailoring to teacher needs (see Becker, Bradshaw et al. [2013] and Becker, Darney et al. [2013] for additional details on the coaching model). Prior research demonstrated that teachers in the integrated condition had more contacts with the coach over the course of the school year (Pas et al., 2015), but that teachers in both conditions implemented a statistically equivalent number of games (Domitrovich et al., 2015).

Measures

    Student measures. At baseline (fall), teachers provided ratings of each student's classroom behaviors over the last three weeks using the *Teacher Observation of Classroom Adaptation-Revised* (i.e., TOCA-R; Werthamer-Larsson, Kellam, & Wheeler, 1991). The TOCA-R includes items on a range of subscales, using a 6-point Likert scale related to behavior frequency (1= *almost never* to 6 = *almost always*) and has been used in several prevention trials (e.g., Ialongo et al., 1999; Petras, Chilcoat, Leaf, Ialongo, & Kellam, 2004; Werthamer-Larsson, Kellam, & Wheeler, 1991). Included in this study were the subscales regarding students' aggressive behavior (15 items; e.g., lied, started physical fights, stubborn, broke rules, hurt others physically, yelled at others; alpha [α = .96]), inattention/hyperactivity (6 items; e.g., paid attention, stays on task, concentrates on class work α = .87), academic engagement (3-items; i.e., completed assignments, learned up to ability, eager to learn; α = .89), and positive peer relations (3 items; i.e., liked by classmates, other children sought him/her out to play, disliked by classmates; α = .83). Two other teacher rating subscales came from the Fast Track Social

Competence Scale: social competence (8 items; e.g., resolves peer problems on his/her own,

expresses feeling appropriately, showed empathy and compassion for others feelings; α = .94)

and emotion regulation (4 items; e.g., controlled temper when there was a disagreement, could

calm down when excited or all wound up, coped well with disappointment or frustration; α

= .88). Items for each subscale were average together to create subscale scores for each student,

ranging from 1 to 6. Higher scores indicated higher frequencies, on average, for the behaviors

captured by each subscale. These subscale scores have been used extensively and there is

considerable data documenting their reliability and validity (e.g., Bradshaw & Kush, 2019;

Ialongo et al., 2018; Werthamer-Larsson et al., 1991; Koth, Bradshaw, & Leaf, 2009).

Teacher beliefs and perceptions. At baseline, teachers completed a self-report survey

regarding their beliefs and perceptions at four time points across the school year. The baseline

data collection for three of the included scales were included in these analyses. Specifically, we

assessed teacher efficacy, burnout, and perceptions of the school organizational climate. The

Behavior Management Self-Efficacy Scale (Main & Hammond, 2008) was a 14-item scale that

included items about classroom behavior management such as "I am able to use a variety of

behavior management techniques" (α = .93). Item response options were based on a 5-point

Likert scale, ranging from "not at all" to "a very great extent", and were averaged, with higher

scores indicating greater efficacy. We also included the *Emotional Exhaustion* scale of the

Maslach Burnout Inventory (Maslach et al. 1997). This included 9 items, such as "I feel used up

at the end of the workday" (α = .91). Responses were rated on a 7-point scale from "never" to

"every day" and averaged to create scale scores such that higher scores indicated greater

emotional exhaustion. Finally, teachers completed the 37-item *Organizational Health Inventory*

for Elementary Schools (OHI; Hoy & Feldman, 1987). Factor analysis has confirmed the

following five-factor structure of the OHI (Hoy & Tarter, 1997): teacher affiliation, academic emphasis, collegial leadership, resource influence, and institutional integrity. The average of all 37 items was calculated to create the OHI scale score ($\alpha = .93$). Item responses were rated on a 4-point Likert scale ranging from "strongly disagree" to "strongly agree". A higher score indicated greater organizational health. These scales were included given their prior use in studies and findings that they are associated with GBG implementation (Domitrovich et al., 2015).

Teacher demographics. Teachers additionally completed a self-report informational form at baseline, providing their own demographic data (i.e., gender and age) and teaching experience data (e.g., grade level taught). Both teacher age and grade taught were included as control variables. Gender was not included because of the high proportion of women in the sample.

PAX GBG implementation dosage. Teachers completed and submitted a weekly log of the number of PAX GBG games played and the duration of each game to the coaches, to document their PAX GBG dosage. These data were summed across the 31-week implementation period. Although not perfectly correlated, there was a high degree of overlap between the two totals (*number of games* and *number of minutes*) and we elected to use the total number of games as our outcome measure. The total count of games ranged in the sample from 1 to 433 with a diffuse distribution across the range, so we considered it quite reasonable to treat the outcome as continuous. We transformed the total counts into z-scores (*ZGP*) to ease interpretation and to avoid an ill-conditioned Level 2 covariance matrix as the range and variance of the totals on the original scale were several orders of magnitude larger (10^4) than the Level 2 predictors..

Overview of the Analyses

To examine whether teacher ratings of student behavior were significantly associated

with teacher implementation levels of the GBG, we conducted a series of multilevel analyses on

the TOCA-R student behavior data in *Mplus,* Version 8.2 (Muthén & Muthén, 1998-2019). At

Level 1, each student-level behavior score for student *i* in classroom *j* was specified as the sum

of a classroom-level (latent) mean (i.e., random intercept), μ_j, plus the student's (latent) deviation

from their classroom mean, ε_{ij}; that is,

$$behavior_{ij} = \mu_j + \varepsilon_{ij}.$$

The random errors at Level 1 for the six behavior subscales (academic achievement, social

competency, positive peer relations, emotion regulation, hyperactivity, and aggression) were

allowed to freely covary with each other. At Level 2, the outcome was specified as a linear

combination of the (latent) classroom-level means for each behavior plus a random disturbance

term; that is,

$$zgp_j = \beta_{ae}\mu_{aej} + \beta_{sc}\mu_{scj} + \beta_{pr}\mu_{prj} + \beta_{er}\mu_{erj} + \beta_{hy}\mu_{hyj} + \beta_{ag}\mu_{agj} + \zeta_j.$$

The random means at Level 2 for the six behaviors were allowed to freely covary with each

other. Figure 1 provides a path diagram representation of the unadjusted, 2-level model.

     After estimating the unadjusted model, we added student gender (grand-mean centered)

as a covariate for all six behavior subscales on the Level 1 model, changing the classroom-level

random intercepts, μ_j, to gender-adjusted latent means. We then added teacher-related covariates

at Level 2 as predictors of the number of Good Behavior games played and correlates of all six

latent classroom-level adjusted behavior subscales mean. We included the behavior management

self-efficacy scale, the emotional exhaustion subscale of the burnout inventory, the

organizational health scale, teacher age (dichotomized to indicate teachers under the age of 30),

and teacher grade level (dichotomized to indicate 3rd-5th grade teachers versus Kindergarten-2nd

grade teachers), some of which have been shown to relate to implementation in prior research

(e.g., Domitrovich et al., 2015). Intervention condition (i.e., PAX GBG vs. PATHs integrated with PAX GBG) was not included, as prior research has indicated that condition was not significantly related to dosage (Domitrovich et al., 2016).

Missing Data. Missing student behavior (i.e., via the TOCA-R) data were addressed through use of full information maximum likelihood estimation. Chi-square analyses revealed no significant differences between the presence of data regarding students based on their gender (χ^2 = 1.246, p = .266).

Results

Descriptive Statistics

Descriptive statistics for the study variables are reported in Tables 1 and 2. There was substantial variation in the number of games played with a mean of 153 games and SD = 100. Overall, the subscales of student behavior were highly correlated with one another in the expected directions at both the within- and between-classroom levels. Similarly, the teacher reported measures of beliefs and perceptions about teaching were correlated with one another in the expected directions.

Implementation Prediction

The results of the final two-level SEM model are presented in Table 3. These are based on the Level 2 regression of the z-score of the total number of Good Behavior games played on the classroom-level latent means of the student behavior subscales, adjusted for gender on the student level (Level 1) and for teacher-related covariates at Level 2. Of the six behavior subscales, the only one statistically significantly related to the outcome was classroom-level aggression. A positive one point difference of mean classroom aggression (1-6 scale) corresponded, on average, to nearly a full standard deviation less of games played (*Est.* = -0.89

[-1.65,-0.12], *S.E.* = 0.39, *p* = .02, *Std. Est.* = -0.33), controlling for all other classroom-level

mean behaviors and teacher-related covariates. Of the teacher-related covariates, only the

emotional exhaustion subscale was significantly related statistically to the outcome. A positive

one point difference in emotional exhaustion corresponded, on average, to a negative difference

of 0.13 standard deviations of games played (*Est.* = -0.13 [-0.24,-0.03], *S.E.* = 0.05, *p* = .02, *Std.*

Est. = -0.18), controlling for all other teacher-related covariates and classroom-level mean

behaviors. The standardized effect estimate of classroom-level aggression was almost double the

standardized effect of emotional exhaustion. There were no variables in the Level 2 model with

a significant positive association with the outcome.

Teacher-rated Student Behaviors and Teacher Self-reported Beliefs and Perceptions

In the final two-level SEM model, teacher-related covariates at Level 2 were allowed to

freely covary with the latent classroom-level means of the student behaviors subscale scores.

The model-estimated correlations are presented in Table 4. Scores for the teacher-reported

behavior management self-efficacy scale had a statistically significant positive association with

classroom-level means of teacher-rated student academic engagement, social competence, peer

relations, and emotion regulation and had a statistically significant negative association with

hyperactivity and aggression. A similar pattern was observed for the organizational health scale.

Emotional exhaustion scores were statistically significantly and negatively associated with

academic engagement and positively associated with aggression.

Discussion

In this study, we sought to build on and extend Dishion and colleagues seminal work on

peer contagion and deviance training in group-delivered interventions into the elementary school

classroom context (see Dishion & Dodge, 2005, 2006; Dodge, Dishion, & Lansford, 2006;

Lochman, Dishion et al., 2017; 2015). More specifically, we examined whether classroom-level student behaviors, both positive and negative, were associated with teachers' implementation of an evidence-based, classroom-wide intervention designed to reduce aggressive/disruptive behavior—the PAX Good Behavior Game (PAX GBG; Embry et al. 2003; Ialongo et al., 2019). We were particularly interested in whether student behavior served as a possible barrier to or as a motivator for high dosage implementation of the program. In the current study, we leveraged a novel and sophisticated analytic approach, whereby student behavioral ratings by teachers were not simply averaged but latently aggregated from the student level to the classroom level.

Our findings suggested that aggressive/disruptive behavior may have served as the primary barrier for high dosage implementation, as teachers reporting a one-point higher level of classroom aggression at baseline, subsequently, played a full standard deviation fewer games across the school year. Although we cannot draw a casual inference based on this association, it is possible that the aggressive/disruptive behavior may make it more challenging for teachers to manage the added burden of implementing the program at a high dosage. This finding is consistent with research on peer contagion in group delivered interventions (Lochman, Dishion, et al., 2017; 2015) and prior work on positive behavioral interventions and supports (Bradshaw, Mitchell, & Leaf, 2009; Pas et al., 2015). Specifically, the program may be perceived as placing an additional demand on the teachers, who are presumably already harried by the high level of problem behavior occurring in the classroom. As such, teachers may struggle to find sufficient time for high implementation dosage. On the other hand, other baseline behaviors (i.e., a range of positive student behavior and hyperactivity) was not associated with implementation dosage. Although we cannot draw causal conclusions based on this null finding, it does suggest that

teachers neither felt motivated nor burdened by the students' baseline positive behaviors when it came to program implementation dosage.

This is among the first studies to examine how positive student behavior relates to implementation of a classroom-based, evidence-based intervention. For example, while student behavior was assessed in a recent study by Sutherland and colleagues (2018b), positive behaviors were excluded. Although one may view negative and positive student behaviors as being two sides of the same coin, the findings from this study suggest that they are differentially associated with dosage. Further, it is possible that hyperactivity is a less challenging behavior than aggressive and disruptive behaviors to teachers, indicating low-level negative behaviors may not be a barrier to implementation. This could indicate there is a "tipping" point at which negative behavior presents a challenge to implementation and should be examined in more depth. Additional research is needed to determine whether there are specific conditions under which implementation dosage can be promoted by student behavior specifically. Further, the extent to which the association between baseline student problem behaviors and reduced dosage serves as a pathway to reduced student outcomes is also an area of future research.

The current study also builds upon earlier research by informing our understanding of the extent to which the relationship between teacher burnout/emotional exhaustion and dosage (Domitrovich et al., 2015) would still hold, while accounting for student behavior. Such inclusion of student behavior has been absent in the prior GBG implementation literature (i.e., Domitrovich et al., 2015; Wehby et al., 2012). Although not a central focus of this study, the fact that the assoications between burnout and dosage remained significant with this additional level of adjustment in the models provided further evidence of the robustness of these associations, which have largely been overlooked in prior research.

Another area of innovation in this study was the use of ML-SEM in order to model the multilevel design of the study. This analytic methodology could be useful for prevention and intervention researchers interested in understanding how individual factors impact implementation of programs at a higher level (e.g., level two, classroom). It allows the researcher to maintain the nuanced and complex nature of a level one (e.g., student) data point, rather than eliminating variability through aggregation. In the absence of conducting a multilevel model, the associations between classroom behavior and intervention implementation in our study may not have been found.

Limitations

Despite several strengths of this study, it is important to consider a number of limitations. For example, this particular study relied on one method of data collection: teacher self report. Teachers provided ratings of the students on the TOCA-R, and prior research suggests that teacher beliefs and perceptions (e.g., organizational factors, interpersonal factors) may actually influence teachers' ratings on the TOCA-R (e.g., see Pas & Bradshaw, 2014). Indeed, this study similarly shows correlations between such teacher beliefs and the student ratings. Teachers also self-reported the number of games played on a weekly basis, as it was not feasible to track the implementation dosage data through another method. Furthermore, prior studies from this trial suggest relatively limited variability in implementation quality of the PAX GBG games, and thus we focused exclusively on dosage, which demonstrated greater variability across the sample (Domitrovich et al., 2015; 2016). Additional variables, such as teacher and student buy-in, may also be of value to explore.

Another limitation is the non-causal nature of this study. For example, teachers are not randomly assigned students with varying levels of student behavior problems. Thus, we cannot

drawa causal conclusions regarding the directionality of associations between student aggressive

and disruptive behavior and teacher implementation dosage or the extent to which it is mediated

by factors, such as perceived burden, stress, motivation, or limited time to implement the

program. Finally, there are limitations which stem from the two-intervention design of the study.

Random assignment to condition occurred at the school level, and there were in fact two

conditions (PAX GBG only vs. integrated PATHs to PAX). Although sensitivity analyses

suggested no differences in the implementation dosage outcome across these two conditions (p

=.81), the lack of power to contrast these findings at a school level represents an additional

potential limitation. We also did not have information on PATHS implementation across both

conditions, and thus were limited to examining just the dosage of the PAX GBG games.

Conclusions and Implications

Taken together, the findings provide compelling evidence that baseline student

aggressive/disruptive behaviors may serve as a potential barrier, rather than a motivator, for

implementation of the PAX GBG. Whereas students in higher aggressive classrooms on average

may have the most potential to actually benefit from preventive interventions, like the PAX

GBG, it appears that high dosage was not achieved. It is possible that additional coaching

support, beyond what was already provided, was needed for teachers working in these settings,

(e.g., more regular, daily check-ins). As such, these classroom-level contextual factors may serve

as an important tailoring variable to consider in future implementations of this and other

behavior management programs.

References

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. (2011). Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research, 20*, 40–49. doi:10.1002/mpr.329.

Barrish, H., Saunders, M., & Wolf, M. (1969). Good Behavior Game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis, 2*, 119–124.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Quantitative Methods in Psychology, 107*, 238-246. doi: 10.1037/0033-2909.107.2.238

Bradshaw, C.P., Koth, C.W., Thornton, L.A., & Leaf, P.J. (2009). Altering school climate through school-wide Positive Behavioral Interventions and Supports: Findings from a group-randomized effectiveness trial. Prevention Science, 10(2), 100-115. doi: 10.1007/s11121-008-0114-9

Bradshaw, C.P., & Kush, J. (2019). TOCA-C: An efficient approach to measuring children's social, emotional, and behavioral functioning by teachers. Manuscript submitted for publication. *Children & Schools.*

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2015). Examining variation in the impact of school-wide positive behavioral interventions and supports: Findings from a randomized controlled effectiveness trial. *Journal of Educational Psychology, 107*(2), 546-557. doi: 10.1037/a0037630.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). New Bury Park, CA: Sage.

Chiapa, A., Parra Morris, G., Véronneau, M. H., & Dishion, T. J. (2016). Translational research on parenting of adolescents: Linking theory to valid observation measures for family centered prevention and treatment. *Translational Behavioral Medicine*, *6*, 90-104.

Dishion, T. J., & Tipsord, J. M. (2011). Peer contagion in child and adolescent social and emotional development. *Annual Review of Psychology, 62,* 189–214. http://dx.doi.org/10.1146/annurev.psych.093008.100412

Dishion, T.J., Nelson, S.E., & Kavanagh, K. (2003). The Family Check-up with high-risk young adolescents: Preventing early-onset substance use by parent monitoring. *Behavior Therapy,* 34, 553–571.doi:http://dx.doi.org.proxy.bibliotheques.uqam.ca:2048/10.1016/S0005-7894(03)80035-7.

Dishion, T. J., & Dodge, K. A. (2005). Peer contagion in interventions for children and adolescents: moving towards an understanding of the ecology and dynamics of change. *Journal of Abnormal Child Psychology, 33*(3), 395-400.

Dishion, T. J., & Dodge, K. A. (2006). Deviant peer contagion in interventions and programs: An ecological framework for understanding influence mechanisms. In K. A. Dodge, T. J. Dishion, & J. E. Lansford (Eds.), *Deviant peer influences in programs for youth: Problems and solutions* (pp. 14–43). New York: Guilford Press.

Dodge, K. A., Dishion, T. J., & Lansford, J. E. (2006). *Deviant peer influences in programs for youth: Problems and solutions.* New York, NY: Guilford Press.

Domitrovich, C. E., Bradshaw, C. P., Greenberg, M. T., Embry, D., Poduska, J. M., & Ialongo, N. S. (2010). Integrated models of school-based prevention: Logic and theory. *Psychology in the Schools, 47*, 71-88. doi: 10.1002/pits.20452

Domitrovich, C., Bradshaw, C.P., Berg, J., Pas, E., Becker, K., Musci, R., Embry, D., & Ialongo, N. (2016). How do school-based prevention programs impact teachers? Findings from a randomized trial of an integrated classroom management and social-emotional program. *Prevention Science, 17,* 325-337.

Domitrovich, C. E., Bradshaw, C. P., Poduska, J. M., Hoagwood, K. E., Buckley, J. A., Olin, S., . . . Ialongo, N. S. (2008). Maximizing the implementation quality of evidence-based preventive interventions in schools: A conceptual framework. *Advances in School Mental Health Promotion, 1*, 6-28. doi: 10.1080/1754730X.2008.9715730

Domitrovich, C. E., Gest, S. D., Gill, S., Jones, D. J., & DeRouise, R. S. (2009). Teacher factors related to the professional development process of the Head Start REDI intervention. *Early Education and Development, 20*, 402–430.

Domitrovich, C., Pas, E. T., Bradshaw, C. P., Becker, K., Keperling, J., Embry, D & Ialongo, N. (2015). Individual and organizational school factors that influence implementation of the PAX Good Behavior Game intervention. *Prevention Science, 16,* 1064-1074. doi: 10.1007/s11121-015-0557-8

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*(3-4), 327-350. doi: 10.1007/s10464-008-9165-0

Embry, D., Staatemeier, G., Richardson, C., Lauger, K., & Mitich, J. (2003). *The PAX Good Behavior Game, First edition*. Center City, MN: Hazelden.

Gottfredson, D. C., Cook, T. D., Gardner, F. E. M., Gorman-Smith, D., Howe, G. W., Sandler, I. M., Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness,

and scale-up research in prevention science: Next generation. *Prevention Science, 16,*

893-926. doi: 10.1007/s11121-015-0555-x

Greenberg, M.T., Kusché, C.A., & Conduct Problems Prevention Research Group (2011).

Grade level PATHS (Grades 3-5). South Deerfield, MA: Channing-Bete Co.

Han, S. S., & Weiss, B. (2005). Sustainability of teacher implementation of school-based mental

health programs. *Journal of Abnormal Child Psychology, 33*(6), 665-679. doi:

10.1007/s10802-005-7646-2

Heilbrun, A., Cornell, D., & Konold, T. (2018). Authoritative school climate and suspension

rates in middle schools: Implications for reducing the racial disparity in school discipline.

Journal of School Violence, 17(3), 324. doi: 10.1080/15388220.2017.1368395

Huang, F. L., & Cornell, D. G. (2016). Using multilevel factor analysis with clustered data:

Investigating the factor structure of the Positive Values Scale. Journal of

Psychoeducational Assessment, 34(1), 3-14.

Ialongo, N., Domitrovich, C., Embry, D., Greenberg, M., Lawson, A., Becker, K., & Bradshaw,

C. (2019, March 21). A randomized controlled trial of the combination of two school-based

universal preventive interventions. *Developmental Psychology*. Advance online publication.

doi: 10.1037/dev0000715.

Ialongo, N., Wethamer, L., Kellam, S., Brown, C., Wang, S., & Lin, Y. (1999). Proximal impact

of two first-grade preventive interventions on the early risk behaviors for later substance

abuse, depression, and anti-social behavior. *American Journal of Community Psychology,*

27, 599–641.

Kam, C., Greenberg, M. T., & Kusché, C. A. (2004). Sustained effects of the PATHS curriculum on the social and psychological adjustment of children in special education. *Journal of Emotional and Behavioral Disorders, 12,* 66-78.

Kam, C.-M., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. *Prevention Science, 4*, 55-63.

Kellam, S. G., Wang, W., Mackenzie, A. C. L., Brown, C. H., Ompad, D. C., Or, F., . . . Windham, A. (2014). The impact of the Good Behavior Game, a universal classroom-based preventive intervention in first and second grades, on high-risk sexual behaviors and drug abuse and dependence disorders into young adulthood. *Prevention Science, 15*, 6-18. doi: 10.1007/s11121-012-0296-z

Kim, E. S., Dedrick, R. F., Cao, C., & Ferron, J. M. (2016). Multilevel factor analysis: Reporting guidelines and a review of reporting practices. *Multivariate Behavioral Tesearch*, *51*(6), 881-898.

Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2009). Teacher Observation of Classroom Adaptation-Checklist: Development and factor structure. *Measurement and Evaluation in Counseling and Development, 42*, 15-30.

Kusché, C.A., Greenberg, M.T., & Conduct Problems Prevention Research Group (2011). Grade level PATHS (Grades 1-2). South Deerfield, MA: Channing-Bete Co.

Latimore, A.D., Burrell, L., Crowne, S., Ojo, K., Cluxton-Keller, F., Gustin, S…. Duggan, A. (2017). Exploring multilevel factors for family engagement in home visiting across two national models. *Prevention Science, 18,* 577-589.doi:10.1007/s11121-017-0767-3.

Lochman, J.E., Dishion, T.J., Boxmeyer, C.L., Powell, N., & Qu, L. (2017). Variation in

    response to evidence-based group preventive intervention for disruptive behavior

    problems: a view from 938 Coping Power sessions. Journal of Abnormal Child

    Psychology, 45, 1271- 1284. https://doi.org/10.1007/s10802-016-0252-7

Lochman, J.E., Dishion, T.J., Powell, N., Boxmeyer, C.L., Qu, L. & Salle, M. (2015).

    Evidence-based preventive intervention for preadolescent aggressive children: One-year

    outcomes following randomization to group versus individual delivery. Journal of

    Consulting and Clinical Psychology, 83, 728-735. http://dx.doi.org/10.1037/ccp0000030

Muthén, B. & Muthén, L. (1997-2019). Mplus User's Guide. Los Angeles, CA: Muthén &

    Muthén.Pas, E., & Bradshaw, C. P. (2014). What affects teacher ratings of student

    behaviors? The potential influence of teachers' perceptions of the school environment

    and experiences. Prevention Science, 15, 940-950. Doi: 10.1007/s11121-013-0432-4

Pas, E. T., Bradshaw, C. P., Becker, K., Domitrovich, C., Berg, J., Musci, R., & Ialongo, N.

    (2015). Trajectories for coaching dosage as a means for improving implementation of the

    Good Behavior Game. School Mental Health, 7, 61-73. doi: 10.1007/s12310-015-9145-0

Pas, E. T., Waasdorp, T. E., & Bradshaw, C. P. (2015). Examining contextual influences on

    classroom-based implementation of positive behavior support strategies: Findings from a

    randomized controlled effectiveness trial. Prevention Science, 16, 1096-1106. doi:

    10.1007/s11121-014-0492-0.

Patterson, G.R., Reid J., & Dishion, T. (1992). A social learning approach: IV. Antisocial boys.

    Eugene, OR: Castalia.

Payne, A. A., Gottfredson, D. C., & Gottfredson, G. D. (2006). School Predictors of the Intensity of Implementation of School-Based Prevention Programs: Results from a National Study. *Prevention Science, 7*, 225-237. doi: 10.1007/s11121-006-0029-2

Pelham, W. E., Dishion, T. J., Tein, J.-Y., Shaw, D. S., & Wilson, M. N. (2017). What doesn't work for whom? Exploring heterogeneity in responsiveness to the Family Check-Up in early childhood using a mixture model approach. *Prevention Science, 18*, 911-922. doi: 10.1007/s11121-017-0805-1

Petras, H., Chilcoat, H. D., Leaf, P. J., Ialongo, N. S., & Kellam, S. G. (2004). Utility of TOCA-R scores during the elementary school years in identifying later violence among adolescent males. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*, 88-96.

Ringwalt, C. L., Ennett, S., Johnson, R., Rohrbach, L. A., Simons-Rudolph, A., Vincus, A., & Thorne, J. (2003). Factors associated with fidelity to substance use prevention curriculum guides in the nation's middle schools. *Health Education & Behavior, 30*, 375-391. doi: 10.1177/1090198103030003010

Roberts, A. M., LoCasale-Crouch, J. DeCoster, J., Hamre, B. K., Downer, J. T., Williford, A. P., Pianta, R. C. (2015). Individual and contextual factors associated with pre-kindergarten teachers' responsiveness to the MyTeachingPartner coaching intervention. *Prevention Science, 16,* 1044-1053. doi: 10.1007/s11121-014-0533-8.

Rohrbach, L. A., Graham, J. W., & Hansen, W. B. (1993). Diffusion of a school-based substance abuse prevention program: Predictors of program implementation. *Preventive Medicine, 22*, 237-260.

Smith, J. D., Berkel, C., Rudo-Stern, J., Montaño, Z., St George, S. M., Prado, G., Mauricio, A. M., Chiapa, A., Bruening, M. M., … Dishion, T. J. (2018). The Family Check-Up 4 Health (FCU4Health): applying implementation science frameworks to the process of adapting an evidence-based parenting program for prevention of pediatric obesity and excess weight gain in primary care. *Frontiers in Public Health*, *6*, 293. doi:10.3389/fpubh.2018.00293

Spoth, R., Rohrbach, L. A., Greenberg, M., Leaf, P., Brown, C. H., Fagan, A., . . . Hawkins, J. D. (2013). Addressing core challenges for the next generation of type 2 translation research and systems: The Translation Science to Population Impact (TSci Impact) Framework. *Prevention Science, 14*(4), 319-351.

Sutherland, K. S., Conroy, M. A., McLeod, B. D., Algina, J., & Kunemund, R. L. (2018a). Factors Associated with Teacher Delivery of a Classroom-Based Tier 2 Prevention Program. *Prevention Science, 19*(2), 186-196. doi: 10.1007/s11121-017-0832-y

Sutherland, K. S., Conroy, M. A., McLeod, B. D. Algina, J., & Wu, E. (2018b). Teacher competence of delivery of BEST in CLASS as a mediator of treatment effects. *School Mental Health, 10,* 214-225. doi: 10.1007/s12310-017-9224-5

Wanless, S. B., Rimm-Kaufman, S. E., Abry, T., Larsen, R. A., & Patton, C. L. (2015). Engagement in Training as a Mechanism to Understanding Fidelity of Implementation of the Responsive Classroom Approach. *Prevention Science, 16,* 1107-1116. doi: 10.1007/s11121-014-0519-6.

Williford, A. P., Sanger Wolcott, C., Vick Whittaker, J. & Locasale-Crouch, J. (2015). Program and teacher characteristics predicting the implementation of Banking Time with preschoolers who display disruptive behaviors. *Prevention Science, 16,* 1054-1063. doi: 10.1007/s11121-015-0544-0.

Wehby, J. H., Maggin, D. M., Moore Partin, T. C., & Robertson, R. (2012). The impact of working alliance, social validity, and teacher burnout on implementation fidelity of the Good Behavior Game. *School Mental Health, 4*(1), 22-33. doi: 10.1007/s12310-011-9067-4

Werthamer-Larsson, L., Kellam, S., & Wheeler, L. (1991). Effect of first grade classroom environment on shy behavior, aggressive behavior, and concentration problems. *American journal of community psychology, 19*, 585-602.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine, 30*, 377–399. doi: 10.1002/sim.4067.

Table 1

Descriptive Statistics for Student-level (Level 1) Variables (n = 3115)

Student-Level Variables	ae	sc	pr	er	hy	ag	M	SD	Min	Max	ICC
Academic Engagement (ae)	1.00	.76	.66	.67	-.72	-.54	4.29	1.24	1.00	6.00	.20
Social Competence (sc)	.68	1.00	.72	.76	-.71	-.53	4.01	1.19	1.00	6.00	.23
Peer Relations (pr)	.61	.74	1.00	.57	-.52	-.52	4.58	1.05	1.00	6.00	.27
Emotion Regulation (er)	.49	.80	.63	1.00	-.69	-.71	4.15	1.24	1.00	6.00	.18
Hyperactivity (hy)	-.73	-.78	-.64	-.71	1.00	.70	2.70	1.07	1.00	6.00	.15
Aggression (ag)	-.48	-.73	-.62	-.80	.72	1.00	1.92	0.94	1.00	6.00	.16
Gender (1-male, 0-female)	-.20	-.24	-.17	-.24	.27	.23	0.50	0.50	0.00	1.00	

Note. Columns 1-6 are correlations; lower diagonal values are the maximum-likelihood estimated Level-1 (within-classroom) correlations and upper diagonal values are the Level-2 (between-classroom) correlations.

Table 2

Descriptive Statistics for Teacher/Classroom-level (Level 2) Variables (N = 204)

Teacher/Classroom-Level Variables	gp	bme	ee	oh	lt30	M	SD	Min	Max
# of Games Played (gp)	1.00					153.18	100.05	1.00	433.00
Behavior Management Efficacy (bme)	.07	1.00				3.85	0.61	1.86	5.00
Emotional Exhaustion (ee)	-.22	-.35	1.00			3.39	1.41	1.00	6.89
Organizational Health (oh)	.13	.24	-.31	1.00		2.80	0.43	1.57	3.84
Young teacher (<30) (lt30)	-.08	-.21	.08	-.18	1.00	0.45	0.50	0.00	1.00
Grade 3-5 (vs. K-2) (g35)	-.11	.02	-.06	.02	-.05	0.40	0.49	0.00	1.00

Note. Columns 1-5 are correlations.

Table 3

Final ML-SEM Level 2 Regression Model Results for the Outcome, Z-score of Number of Games Played

Level 2 Regression	*Est.*	*S.E.*	*p*	*95% CI*		*Std. Est.*
Academic Engagement (ae)	0.06	0.32	.86	(-0.57 ,	0.68)	0.03
Social Competence (sc)	0.20	0.35	.57	(-0.48 ,	0.88)	0.12
Peer Relations (pr)	0.01	0.25	.97	(-0.47 ,	0.49)	0.01
Emotion Regulation (er)	-0.38	0.37	.30	(-1.11 ,	0.34)	-0.20
Hyperactivity (hy)	0.46	0.38	.22	(-0.28 ,	1.20)	0.19
Aggression (ag)	**-0.89**	**0.39**	**.02**	**(-1.65 ,**	**-0.12)**	**-0.33**
Behavior Management Efficacy (bme)	-0.09	0.15	.54	(-0.39 ,	0.21)	-0.06
Emotional Exhaustion (ee)	**-0.13**	**0.05**	**.02**	**(-0.24 ,**	**-0.03)**	**-0.18**
Organizational Health (oh)	0.09	0.22	.68	(-0.34 ,	0.51)	0.04
Young teacher (<30) (lt30)	-0.18	0.16	.27	(-0.49 ,	0.14)	-0.09
Grade 3-5 (vs. K-2) (g35)	-0.26	0.17	.12	(-0.59 ,	0.07)	-0.13

Table 4

Model-estimated Correlations between Teacher-Level Covariates and Classroom-Level Latent Behavior Subscale Means

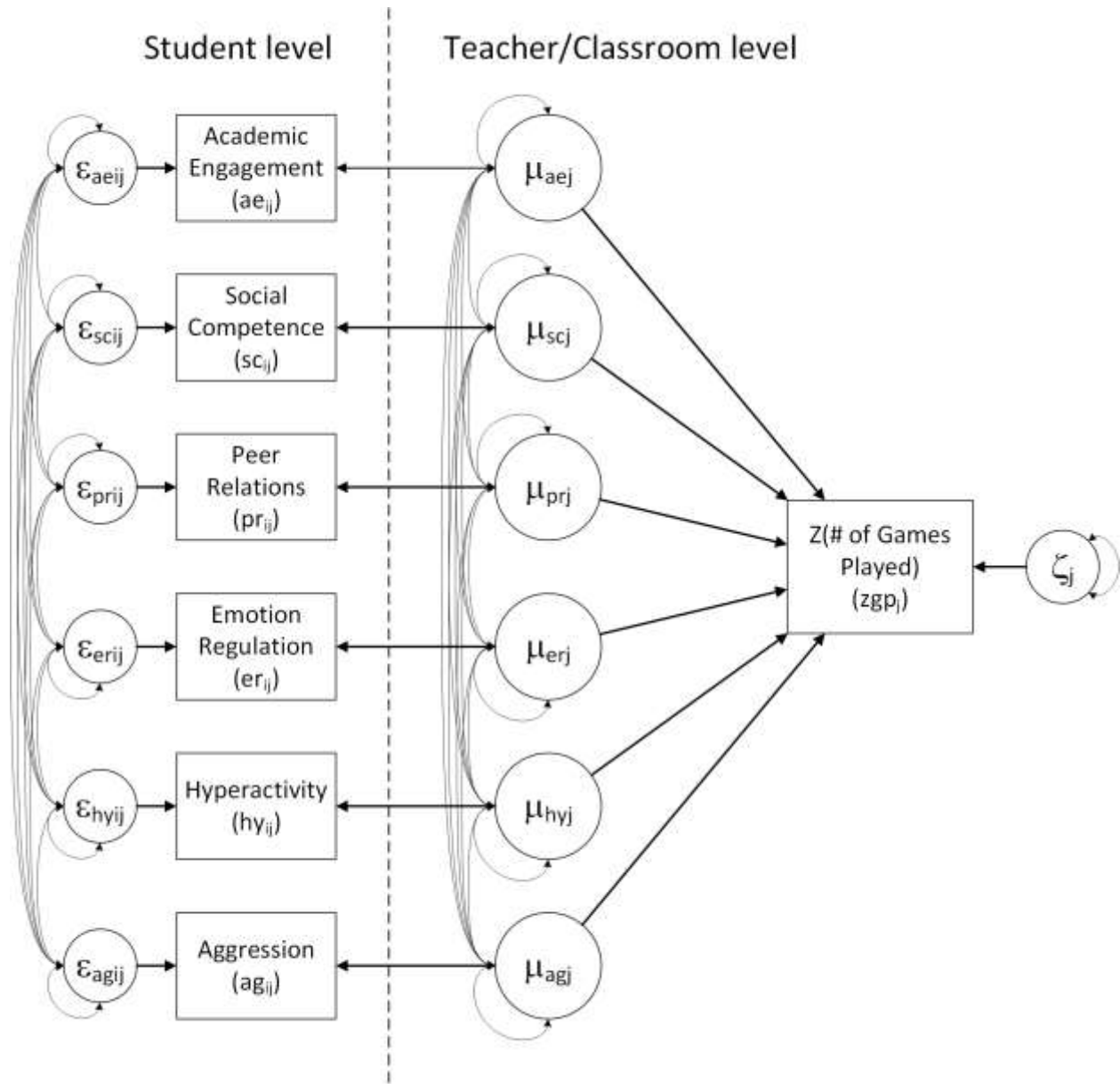|	Academic Engagement (ae)	Social Competence (sc)	Peer Relations (pr)	Emotion Regulation (er)	Hyperactivity (hy)	Aggression (ag)
Behavior Management Efficacy (bme)	**.31**	**.23**	**.24**	**.22**	**-.32**	**-.32**
Emotional Exhaustion (ee)	**-.17**	-.08	-.14	-.05	.10	**.18**
Organizational Health (oh)	**.34**	**.30**	**.24**	.14	**-.31**	**-.23**
Young teacher (<30) (lt30)	-.02	-.06	**-.18**	-.01	.09	.01
Grade 3-5 (vs. K-2) (g35)	**-.25**	-.05	**-.19**	**-.18**	.05	.07

Note. **Bold**: $p < .05$.

Figure 1. Path diagram of the unadjusted multilevel SEM predicting the teacher-reported number of games played by the latent classroom-level means of six student behavior subscales.