Individual and School Organizational Factors that Influence Implementation of the PAX Good Behavior Game Intervention

Celene E. Domitrovich[1]

Elise T. Pas[2]

Catherine P. Bradshaw[2]

Kimberly D. Becker[3]

Jennifer Keperling[2]

Dennis Embry[4]

Nicholas Ialongo[2]

Pennsylvania State University[1]

Johns Hopkins University, Bloomberg School of Public Health[2]

University of Maryland, School of Medicine[3]

Paxis Institute[4]

Domitrovich, C. E., Pas, E. T., Bradshaw, C. P., Becker, K. D., Keperling, J. P., Embry, D. D., & Ialongo, N. (2015). Individual and School Organizational Factors that Influence Implementation of the PAX Good Behavior Game Intervention. *Prevention Science*, *16*(8), 1064–1074. https://doi.org/10.1007/s11121-015-0557-8

Published in *Prevention Science*

Funding: This research was supported in part by grants from the Institute of Education [R305A130060 and R305A080326] and the National Institute of Mental Health [P30 MH08643].

Abstract

Evidence-based interventions are being disseminated broadly in schools across the U.S., but the implementation levels achieved in community settings vary considerably. The current study examined the extent to which teacher and school factors were associated with implementation dosage and quality of the PAX Good Behavior Game (PAX GBG), a universal classroom-based preventive intervention designed to improve student social-emotional competence and behavior. Specifically, dosage (i.e., number of games and duration of games) across the school year and quality (i.e., how well the game is delivered) of PAX GBG implementation across four time points in a school year were examined. Hierarchical linear modeling was used to examine the association between teacher-level factors (e.g., demographics, self-reports of personal resources, attitudes toward the intervention, and workplace perceptions) and longitudinal implementation data. We also accounted for school-level factors, including demographic characteristics of the students and ratings of the schools' organizational health. Findings indicated that only a few teacher-level factors were significantly related to variation in implementation, whereby perceptions (e.g., fit with teaching style, emotional exhaustion) were generally related to dosage, whereas demographic factors (e.g., teachers' age) were related to quality. These findings highlight the importance of school contextual and proximal teacher factors on the implementation of classroom-based programs.

Individual and School Organizational Factors that Influence Implementation of the PAX Good Behavior Game Intervention

The public health approach for improving children's behavioral and educational outcomes in the field of education and prevention science includes a phased process, starting with developing interventions and programs, testing them through randomized controlled trials (RCTs), and then offering them to community organizations or agencies (Flay et al., 2005). This approach has led to the implicit expectation that districts or schools would adopt and implement evidence-based programs with a high degree of adherence to the original model; however, implementation in real-world settings is typically not up to the level achieved in efficacy trials (Durlak & Dupree, 2008). Studying the gap between efficacy and effectiveness in school-based prevention is important because student outcomes have been linked to the quality of program implementation (Durlak & Dupree, 2008). The purpose of the current study was to examine factors associated with implementation of a universal classroom-based preventive intervention designed to improve student behavior, called the PAX Good Behavior Game (PAX GBG; Embry, Staatemeier, Richardson, Lauger, & Mitich, 2003).

Implementation is typically defined as the discrepancy between what is planned and what is delivered (Dane & Schneider, 1998). Operational definitions of implementation include indicators such as dosage (i.e., how much of an intervention is delivered) and quality (i.e., how well and comprehensively an intervention is delivered) in both the intervention and the support system (Chen, 1998). Several conceptual models have been developed to describe the factors that influence implementation in real world settings. For example, the Interactive Systems Framework (ISF) for Implementation and Dissemination highlighted the importance of general capacity (e.g., leadership, climate, organizational innovativeness) and innovation-specific

capacity (e.g., implementation support for the intervention) within the setting where an innovation is being introduced (Wandersman et al., 2008). Han and Weiss (2005) suggested that the process of delivering interventions in educational settings is influenced by organizational features of schools and the quality of a school's environment, as well as characteristics of the individuals delivering the intervention. In their model, personal resources (i.e., efficacy) and positive perceptions of interventions are suggested targets of training and consultation because they are factors that likely foster higher levels of implementation by teachers.

Factors that Influence Implementation Outcomes

Reviews of the empirical literature support these conceptual models and suggest that an even broader set of individual and organizational factors may contribute to variation in how much and how well school-based interventions are conducted (Berkel, Mauricio, Schoenfelder, & Sandler, 2011; Domitrovich et al., 2008; Durlak & Dupree, 2008). Below we consider some of these variables.

Demographic and Professional Characteristics. Teacher demographic and professional characteristics are potentially relevant to patterns of implementation behavior (Rohrbach, Grana, Sussman, & Valente, 2006). Yet these factors have rarely been empirically examined, and the available research is somewhat mixed. For example, one study of middle school preventive interventions found that teacher experience was not associated with implementation (Ringwalt et al., 2003), whereas studies of preschool educators, where there is more variability in the workforce, the findings are inconsistent. Some studies suggest higher education levels are associated with better quality teaching and implementation of a curricular intervention (Domitrovich, Gest, Gill, Jones, & DeRouise, 2009; Tout, Zaslow, & Berry, 2006) while others

fail to show an association between teacher professional characteristics and implementation (Baker, Kupersmidt, Voegler-Lee, Arnold, & Willoughby, 2010).

Personal Resources. Given their influence over motivation and beliefs, researchers have also examined how personal resources such as efficacy and burnout impact teachers' implementation of evidence-based interventions. Self-efficacy has been positively associated with prevention program implementation (Ringwalt et al., 2003; Rohrbach et al., 1993), whereas emotional exhaustion appears to undermine teachers' attitudes towards interventions and implementation quality (Ransford, Greenberg, Domitrovich, Small, & Jacobson, 2009). Yet, a study by Domitrovich et al. (2009) found that emotional exhaustion was positively related to implementation; this may be a context-specific finding as this study was conducted in preschools whereas other studies were conducted in elementary schools.

Workplace Perceptions. Implementation research has demonstrated that the organizational health of schools influences implementation by teachers (Beets et al., 2008; Gottfredson & Gottfredson, 2002; Gregory, Henry, & Schoeny, 2007). Organizational health is reflected in the behavior of school leaders, staff, and students. Principals set expectations for staff behavior and provide support in a variety of ways including how they allocate resources and how they resolve problems that arise. Research has established that staff perceptions of these leadership qualities influence their implementation of interventions (Kam et al., 2003; Payne, Gottfredson, & Gottfredson, 2006; Rohrbach et al., 1993). In addition to having effective leaders, healthy organizations have members who report feeling connected to one another (Beets et al., 2008; Hoy & Feldman, 1987). Finally, schools that are high functioning have high expectations for students, are interested in new ideas, and are open to innovation (Bryk & Schneider, 2002). It is likely that ratings of the workplace made by individual teachers reflect the *true* characteristics

of the organization, as well as *error* that is a function of bias inherent in an individual's perspective and measurement error (Crocker & Algina, 1986). Therefore, ratings made by multiple members of an organization are often averaged in order to measure organizational characteristics such as health or climate that represent the actual quality of a setting as opposed to one individual's perception of the setting. Research has shown that both individual and shared perceptions of the workplace environment are important and have the potential to distinctly influence individual behavior, therefore, in the current study we consider both sets of ratings (Choi et al., 2003).

Intervention Acceptance. Teacher acceptance of an intervention has the potential to influence implementation motivation and behavioral change. According to Han and Weiss (2005), positive attitudes toward an intervention and implementation motivation may initially be dependent on consistency of the program with the teacher's philosophy or current practice. Over time, teacher attitudes may be more dependent on experience with a program and the extent to which it is perceived as effective by the user (Han & Weiss, 2005). In empirical studies, intervention acceptance is associated with higher levels of dosage and quality (Beets et al., 2008; Domitrovich et al., 2009; Ringwalt et al., 2003; Rohrbach et al., 1993; Wehby et al., 2011) whereas negative perceptions are associated with lower levels of implementation (Gottfredson & Gottfredson, 2002).

Implementation Support. The use of coaching as a support strategy to facilitate implementation is growing (Pas, Bradshaw, & Cash, 2014). In one study of the GBG, the quality of the alliance a teacher had with the coach was the strongest predictor of implementation (Wehby, Maggin, Partin, & Robertson, 2011). Coaches have the potential to buffer the effects of stress and help teachers develop efficacy. In the study by Wehby et al. (2011), the quality of

teachers' relationship with the coach also moderated the effect of emotional exhaustion on implementation; this suggests that the extent to which the lack of personal resources undermines implementation may depend on coaching supports.

School Characteristics. Research has shown that schools' structural (e.g., large school size) and student demographic composition (e.g., high concentration of poverty) relate to negative student outcomes and act as proxies for school disorder that could relate to implementation (e.g., Bradshaw, Sawyer, & O'Brennan, 2009; Gottfredson & Gottfredson, 2002). Intervention studies suggest that larger, urban schools with a high concentration of student poverty implement school-based interventions more extensively (Payne et al., 2006). This finding does not speak to the quality with which interventions are implemented but may explain patterns of program use (i.e., dosage).

Current Study

Guided by the work of Han and Weiss (2005) and the ISF conceptual model (Scaccia et al., 2014; Wandersman et al., 2008), the current study examined how a comprehensive set of individual- and school-level factors were associated with teachers' dosage and quality of PAX GBG delivery. This research is important given increasing interest in GBG, as much of the prior work has focused on outcomes (see Ialongo et al., 1999; 2001; Kellam et al., 2008), rather than implementation. Data were drawn from a one-year RCT, testing PAX GBG as implemented both as a stand-alone intervention and as part of a more comprehensive intervention where it was integrated with the Promoting Alternative Thinking Strategies (PATHS) curriculum (Greenberg, Kusché, & CPPRG, 2011; Kusché, Greenberg, & CPPRG, 2011), which is a universal social-emotional intervention. The integration of multiple evidence-based interventions has been proposed in the prevention literature as a strategy to maximize impact and address multiple

mechanisms that influence student outcomes (Domitrovich et al., 2010; Flay et al., 2005). One often-raised concern, however, is that integrated interventions have the potential to be more burdensome to implement and therefore are more likely to be poorly executed (Domitrovich et al., 2010). Given the unique design of the current study, we explored this issue by contrasting the two treatment conditions. This study also builds on previous research by including a more diverse set of variables than typically available in school-based implementation studies and by examining a variety of factors simultaneously using a multi-level approach. Given the lack of previous research on teachers' demographic (i.e., age, gender) or professional characteristics (i.e., education, grade level taught), hypotheses about how these factors would influence the implementation of the GBG were not proposed. Based on the Han and Weiss (2002) model, teachers with higher efficacy regarding the management of student behavior and lower levels of emotional exhaustion were expected to implement with greater quality and dosage compared to those with fewer personal resources and more negative perceptions. Teachers with more positive perceptions of the intervention (i.e., acceptance and motivation) and of the support they received from coaches and administrators were expected to implement at higher levels compared to those with more negative perceptions. Positive school climate, both from an individual perspective and collectively among teachers, was expected to promote higher levels of GBG implementation in terms of both dosage and quality. Implementation levels between the two intervention conditions (i.e., PAX GBG and PAX GBG integrated with PATHS) were compared in order to determine whether the inclusion of PATHS undermined the quality and dosage of GBG. We anticipated comparable quality ratings for teachers in both conditions due to the simplicity of the game and the provision of coaching support, but lower dosage levels when teachers delivered the GBG in combination with PATHS as a function of increased intervention burden.

Method

Design Overview

Data for this study were drawn from an RCT of the integration of the PAX GBG with the PATHS curriculum (Greenberg, Kusché, & CPRG, 2011; Kusché, Greenberg, & CPPRG, 2011). Participating schools were randomized to one of three conditions: the integrated model (9 schools), the PAX GBG (9 schools), and a control condition (9 schools) where teachers conducted their usual practice. The current study included the 18 schools assigned to the two intervention conditions. The schools were all located in a large urban, east coast public school district (see school demographics in Table 1). The students within these schools were predominantly African American (on average, 88%) and the majority received free and reduced meals (i.e., FARMs; 85%). Recruitment 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 in the data collection was voluntary.

Participants

The current study sample included the 222 intervention teachers (K-5) who were recruited across the 18 intervention schools; 108 teachers (54.3%) in this study were in schools randomly assigned to deliver PAX GBG, whereas 91 (45.7%) were in schools randomized to deliver the integrated PAX GBG and PATHS curriculum. The vast majority of sample teachers were female (i.e., 89%) and about half were 30 or younger, had a graduate degree, and taught students in grades 3 through 5. See Table 1 for further details on the sample as well as average scores on the variables measured in this study.

Interventions

Across both intervention conditions, teachers were trained to implement the PAX GBG. Originally developed by Barrish et al. (1969), GBG encourages teachers to utilize social learning principles within a team-based, game-like context to reduce aggressive, disruptive, and off-task behavior and thus facilitate academic instruction. PAX GBG represents Embry and colleagues' (2003) efforts to improve the effectiveness of GBG by primarily adding verbal and visual cues that teachers and students use to promote attentive and prosocial behaviors and a positive classroom environment when not "playing the Game". GBG has positive academic, behavioral, and substance use outcomes (for details see Ialongo et al., 1999; 2001; Kellam et al., 2008).

Teachers in the integrated condition (see Domitrovich et al., 2010 for description) implemented both PATHS (Greenberg, Kusché, & CPRG, 2011; Kusché, Greenberg, & CPPRG, 2011) and PAX GBG. PATHS is a universal social-emotional classroom curriculum that includes a developmentally appropriate series of lessons and activities that provide direct instruction and practice opportunities to develop students' social-emotional skills. Prior RCTs have shown that PATHS reduces off-task, aggressive, and disruptive behaviors by improving prosocial cognitions and socially competent behaviors (Greenberg & Kusche, 2006).

Measures

Quality of PAX GBG Delivery. Rubric ratings of the quality with which teachers delivered the GBG game were completed by coaches and other research staff at four time points throughout the academic year. Members of the team co-observed the first 15% of teachers together at each round of data collection. After establishing reliability at or above ICCs of .80 for each item, the remaining observations of quality were collected independently. Coaches and research staff were randomly assigned to schools each round. Coach ratings of teachers were only used for teachers that were not assigned to the coach. During the observation, teachers were

asked to conduct a 5 to 10-minute game. The *Game Observation* scale of the PAX GBG rubric (Schaffer, Rouiller et al., 2006) includes 7 items which assess the quality with which teachers prepared and executed the game (α = .93). This includes: 1) getting students set up for the game, 2) the activity during which the game is conducted, 3) use of the timer, 4) structure of teams, 5) how behavior is responded to during the game, 6) review of the game after it ends, and 7) the prize that is given after the game. Ratings were made on a scale of 0 to 4, with higher scores indicating better quality implementation. The overall average rating for game quality was 3.28 (SD =.63) and means ranged from 3.11 to 3.34 (SD = .52 to .71) across the four time points.

Dosage of PAX GBG. Teachers completed and submitted a weekly log of the number of PAX GBG games played and the duration of each game to the coaches. These data were summed across the 31-week implementation period and yielded two outcome variables: *total number of games* implemented and *total number of minutes* implementing the PAX GBG.

Demograhic and Professional Characteristics. A teacher information form was used to collect information on teacher demographics (i.e., gender, age, education, years teaching, degrees), professional development experiences, and information regarding other social-emotional and classroom management interventions being used by the teacher. Teachers' gender, age, graduate degree attainment, and the grade level taught were included in the current study.

Personal Resources. The *Behavior Management Self-Efficacy Scale* (Main & Hammond, 2008) has 14 items specific to promoting classroom behavior management (e.g., "I am able to use a variety of behavior management techniques"; α = .94). Item responses were provided on a 5-point Likert-type scale ranging from "*not at all*" to "*a lot*." We also administered the *Emotional Exhaustion* (9 items, e.g., I feel used up at the end of the workday, α = .92) scale of

the teacher report version of the *Maslach Burnout Inventory* (MBI; Maslach et al., 1997). Responses were rated on a 7-point scale from never to every day.

Intervention Acceptance. Teachers' intervention acceptance was assessed with ratings of the extent to which teachers perceived the GBG as fitting with their schedule and teaching style. Specifically, the *Teacher Perceptions of the Intervention Attributes* (TPIA; Domitrovich & Ialongo, 2008a) measure assessed a range of teacher perceptions regarding the program; specifically, two items regarding the extent to which the GBG basics and the PAX cues *Fit with Schedule* (α = .68; 5-point scale ranging from difficult to very easy). Two items regarding how these same elements *Fit with Style* (α = .76; 5-point scale ranging from not at all to a lot) were provided and an average score was calculated for each scale. A single, 5-point Likert-type item from the TPIA reflecting teachers' report of *motivation to implement* was also included as an indicator of intervention acceptance.

Workplace Perceptions. Teachers completed the 37-item *Organizational Health Inventory for Elementary Schools* (OHI; Hoy & Feldman, 1987) to assess the organizational health of their school. Item responses include a 4-point Likert-type scale ranging from "*rarely occurs*" to "*very frequently occurs*." The OHI has been used extensively in educational research. A factor analysis has confirmed the following four-factor structure (Hoy & Tarter, 1997): *teacher affiliation* (9 items, α = .87), *academic emphasis* (5 items, α = .67), *collegial leadership* (10 items, α = .93), and *resource influence* (7 items, α = .87) but studies have combined the scales to create an overall climate indicator (Bevans, Bradshaw, Miech, & Leaf, 2007). This study also included six items related to *institutional integrity*. The total score for the OHI was calculated (α = .93 in this study) by averaging all items; exploratory factor analyses indicated that all items could be retained. In addition, the *Openness to Innovation* scale from the Trust in

Schools measure (Bryk & Schneider, 2002) assessed teachers' perceptions of the school climate in terms of staff attitudes towards innovation. The scale was comprised of three items (e.g., "Teachers and other professional staff are continually learning and seeking new ideas"). Responses to these items were rated on a 4-point Likert-type scale including strongly disagree to strongly agree. A total score ($\alpha = .84$) was created for each teacher by averaging their responses across the three items. The scores for all teachers within a school were also averaged to represent a school-level rating of organizational health.

Implementation Support. Teachers completed the *Teacher-Coach Alliance Scale* (Domitrovich, Poduska, & Bradshaw, 2008), which was adapted from Wehby et al. (2011) in order to measure teachers' perceptions of their working relationships with their coaches. The measure includes 23 items (e.g., "I feel confident in my coach's ability to help me implement PATHS to PAX", "the time I spend working with my coach is effective and productive") that are rated on a 5-point scale (never to always). A total score ($\alpha = .97$) was created by averaging all of the items. Teachers' perceptions of the support provided by their principal for the PATHS to PAX program was assessed with the *Perceptions of Administrative Support Scale* (PASS; Domitrovich & Ialongo, 2008b). The PASS is a 9-item scale regarding the extent to which teachers perceive administrators as modeling and promoting the language and practices of PATHS to PAX. Items are rated on a 4-point scale from "*strongly disagree*" to "*strongly agree*." The scale score ($\alpha = .94$) was created by averaging the items.

School Characteristics. The school size and percent of students receiving FARMs were retrieved from data made publicly available by the Maryland State Department of Education.

Analyses

Missing Data. Though data were collected from all targeted teachers, there were missing data on the various variables included. Specifically, the included variables had missing data on 0 to 20% of cases and thus we used imputation to retain all cases. Data were assumed to be missing at random (MAR) as analyses did not indicate differential missingness based on demographic data. Therefore we used the MICE (multivariate imputation by chained equations) method of multiple multivariate imputation in STATA (Azur, Stuart, Frangkis, & Leaf, 2011; White, Royston & Wood, 2011). MICE imputes each variable conditional on all of the other variables in the imputation procedure, and iterates that process until convergence. Additionally, three interaction terms with teacher related variables were included to account for condition (grade taught, years of experience, and graduate degree). All variables included in the analyses in this paper were included in the imputation procedure. In addition, school level predictors such as school size (enrollment), FARMs, and mobility were included to inform the imputation.

Outcome Analyses. A series of correlational analyses were completed in SPSS to determine that teacher- and school-level variables were not too highly correlated, and therefore would not result in collinearity in the HLM models. Hierarchical linear modeling (HLM) was conducted in HLM 7.01 (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011) to account for nesting of teachers within schools for the three outcomes: the total number of minutes played, the total number of games, and the ratings of game quality given by coaches during observations. The first two measures were assessed at one time point (i.e., each an average for the year) whereas the PAX GBG observations were conducted four times during the school year; therefore, the models that included these indicators did not include a longitudinal (third) level whereas the observations did. For both analyses, the same set of predictor variables was incorporated. The HLM models were built one variable and level at a time to ensure that changes

in the magnitude and direction of effects did not occur; this approach is also recommended to detect collinearity (Raudenbush & Bryk, 2002). We ensured that the findings from final model were consistent with those models where each predictor was modeled individually.

At the teacher-level, the models included teacher demographics (i.e., gender, age), education (i.e., attainment of a graduate degree), grade level taught, personal resources (i.e., efficacy for behavioral management, emotional exhaustion), attitudes toward the intervention (i.e., fit with the schedule, fit with style, self-reported motivation to implement) and the support system (i.e., principal support, quality of coach relationship), and perceptions of the workplace (i.e., school organizational health and openness to innovation). Teacher demographic data, training, and grade level taught were dichotomized (1= female, 0 = male for gender; 1 = 20-30 years old, 0 = 31+ for age; 1 = graduate degree, 0 = no graduate degree; and 1 = 3rd-5th, 0 = K-2nd for grade taught); each of these variables was uncentered. All other teacher-level variables were centered at the grand mean (Luke, 2004). At the school-level, the number of students enrolled, percent of students receiving free and reduced meals (FARMs), and the aggregated school-level OHI score were included. Each variable was centered at the grand mean (Luke, 2004). Treatment status (i.e., PAX GBG only versus integrated) was uncentered.

The number of minutes and number of games variables were count variables and therefore HLM was conducted using the Poisson distribution, which accounts for the fact that these outcome scores are bounded at 0 on the low end and are unbounded on the high end (Cameron & Trivedi, 1998). Because the variance of this outcome exceeded the mean, the assumption that the mean and variance both equal λ is violated (Land, McCall, & Nagin, 1996); therefore, the over-dispersion function in HLM was utilized to adjust the standard errors (Raudenbush et al., 2011). These were two-level models where teacher-level variables were

modeled at Level-1 and school-level variables were modeled at Level-2. Game observations were measured by one continuous variable, examined using a 3-level HLM model with the repeated measures modeled at Level-1, teacher variables at Level-2, and school variables at Level-3. The intercept for this model was set at the baseline measure of dosage.

Results

Number of PAX GBG games played

On average, across the school year, teachers played approximately 152 PAX GBG games (*SD* = 99.74). Two of the teacher-level variables were significantly related to the number of PAX GBG games played: teacher-reported emotional exhaustion and the teachers' perceptions that the games fit with their teacher style. Specifically, teachers who provided higher scores of emotional exhaustion-related burnout reported playing fewer games (γ_{60} = -0.10, event rate ratio [ERR] = 0.91, p = .02). On the other hand, those teachers that reported that GBG fit with their style played more games (γ_{130} = 0.17, ERR = 1.18, p = .02). On average, teachers that provided burnout ratings that were one point higher than the sample average played about 14 fewer games (i.e., almost one-fifth of a standard deviation). Yet, those who provided one point higher scores (than average) on the perceptions of fit with style, played about 27 more games (again, about one-fifth of a standard deviation more games). School-level variables were not significantly related to the number of games played. Contrary to our hypothesis, treatment status was not significantly related to dosage; teachers implementing the integrated PATHS and PAX GBG intervention spent the same amount of time implementing the PAX GBG games as teachers in the PAX GBG only condition. See Table 2 for a full reporting of the final model results.

Number of Minutes of PAX GBG Games Played

On average, teachers spent nearly 1488 minutes (or 24.8 hours) playing the PAX GBG games (SD = 7388.00) across the school year. None of the modeled teacher variables had a statistically significant relationship with this outcome, although ratings of motivation and fit with style both approached significance at the $p < .10$ level. None of the school-level variables were significantly related to the number of games. Again, the number of minutes of PAX GBG was statistically equivalent for teachers in both the integrated PATHS and PAX GBG only conditions. See Table 2 for a full reporting of the final model results.

Quality of PAX GBG Delivery

Young (i.e., ages 20-30) teachers had quality scores that were about one-third of a standard deviation higher at the intercept than older teachers (i.e., 31 and older; $\beta_{030} = 0.22$, $p = .04$). No teacher-level variables were significantly related to either the intercept score or the slope of time. At the school-level, none of the modeled variables were significantly related to implementation quality. A significant slope factor for time indicated that scores increased significantly when no predictors were included in the model, however the significance of this factor diminished in the final model (see Table 3).

Correlations among Outcomes

The two dosage indicators were highly correlated ($r = .74$). However, the dosage variables were not as highly correlated with quality (*rs* ranged .21 to .31 for the dosage indicators and rubric scores at the four time points). This suggests that teachers with higher rubric scores did not necessarily implement the PAX game more frequently or for longer periods of time.

Discussion

The current study examined the associations of dosage and quality of the PAX GBG with individual and school-level factors theoretically linked to these implementation outcomes. Only a

few of the individual factors emerged as significant. This suggests that for a classroom-based intervention such as the PAX GBG which is delivered primarily by teachers to their specific classroom, the characteristics of the building may be less relevant than the characteristics of the teachers within the building. Interestingly, the factors that emerged as significantly associated with PAX GBG implementation outcomes in the current study were primarily related to game dosage as opposed to quality. Teachers' personal resources and attitudes played a role in the frequency with which they played the game but not how long or how well they played. Teachers who reported feeling more exhausted and overwhelmed were less inclined to play the games (i.e., one point above average equated to an average of 138 games versus the average 152) compared to teachers with average ratings of burnout. For teachers who were stressed, initiating a game was likely more burdensome than extending the time spent playing games that were already underway. On the other hand, teachers who felt that the game fit with their teaching style played more games (i.e., about 179 games versus 152 on average) than teachers who did not have this attitude. Alignment between an intervention and its user is often cited as an important criterion for adopting an evidence-based intervention (e.g., Han & Weiss, 2005), but our results suggest that "fit" may also be relevant for how an intervention is used over time. Once an intervention is selected by a school it may be worthwhile for coaches or individuals who support teachers to identify the ways in which the intervention aligns with teachers' philosophies or approaches to learning as a strategy to promote continued use of the intervention.

The fact that the current study included multiple indicators of implementation raises an important question about implementation measurement and the nature of different interventions. Given the relative simplicity of the PAX GBG compared to other classroom-based interventions, ratings of game quality may be less predictive of improvements in student outcomes than

measures of dosage. GBG is unique in the fact that dosage can be measured relative to both the number of games played and the length of each game. Teachers may play the game more frequently when their students are disruptive but they may also chose to play more often because their students enjoy the game and it promotes a positive climate in the classroom. The implementation guidelines provided with the PAX GBG suggest that as students' ability to self-regulate improves, teachers should extend the length of the game. Moreover, games may need to be longer for older students compared to younger students, given developmental improvements in students' capacity to manage their behavior. Without research linking various implementation indicators to student outcomes it is difficult to determine the relative importance of each one. GBG has received a great deal of attention in the past few years which has contributed to its wide disseminated. The expansion of its use should provide more opportunities for further research including research with larger samples and allowing for multi-group analyses.

In the current study, younger teachers had significantly higher PAX GBG quality, but only at the intercept. Discussions with the intervention coaches revealed that some older teachers were less interested in changing their regular teaching practices to incorporate the game. It is possible that younger teachers were more open to trying new management methods and this resulted in higher quality of PAX GBG implementation. It is also possible that because younger teachers have less experience in the classroom, they are more conscientious about following the implementation "rules," particularly when an outside observer was present, whereas older teachers may be more confident in their own abilities, therefore making adaptations that resulted in lower quality ratings. Teacher age was the only individual factor related to implementation.

The average quality scores across the study sample were relatively high (M = 3.14, on a 0-4 scale), so it is possible that the restricted variability in the ratings limited the number of

associations that could be detected with the individual and school-level factors. Achieving a high quality score may also be less dependent on individual differences among teachers. There were seven game delivery elements assessed on the implementation rubric, which are aligned with training and ongoing coach feedback. The repetition of these elements may have made it easy for teachers to conduct the game well in the context of the classroom observation. High quality delivery during this classroom observation may or may not have reflected the degree to which teachers were delivering the game well under natural conditions. In fact, the correlations between quality and dosage, though positive, were small further supporting this hypothesis.

Teachers in the current study were participating in an RCT designed to test whether the integrated PATHS and PAX GBG has a greater impact on student outcomes than the PAX GBG implemented alone. Given this design, it was important to also explore whether implementation of PAX GBG differed between the conditions. Importantly, we found that the amount of time spent implementing PAX GBG and the quality of implementation were statistically equivalent in both conditions, despite the added components of the integrated intervention suggesting that integrated programs can achieve the same level of implementation as non-integrated versions if aligned well (Domitrovich et al., 2010).

This preliminary finding is promising but needs to be replicated and explored further using multi-group analyses. It may be that the nature of the interventions being integrated plays a role in whether or not implementation quality is impacted. The PAX GBG and PATHS work well together because one provides a supportive context for the other and they do not compete for the same classroom time. Other intervention combinations may prove more burdensome. It is also important to recognize that teachers in the current study were supported with weekly

coaching so it is difficult to determine how coaching contributed to the implementation success of one program over the other.

Study Limitations

Despite the strengths of the current study, some limitations should be noted. First, although the final model demonstrated a reduction in variability in the outcomes, there remained variability in the outcomes. This suggests a need for more research in order to fully understand the process of intervention implementation in schools. This could include a variety of approaches. For example, we only examined initial scores on the teacher perceptual variables without consideration of change in these variables over time in response to the impacts of the intervention and/or coaching, as suggested by Han and Weiss (2005). This transactional process requires more careful consideration, perhaps using a parallel process model. There may be additional factors that were not measured in the current study or different implementation indicators that should be assessed. A more direct measure of implementation or of teachers' reasons for implementing an intervention could expand our knowledge on this topic. The current study took place in one large, urban school district and thus do not know the effects of these variables in tribal, rural, or suburban schools. Lastly, our sample size was relatively small for examining school-level factors, given the small number of schools in the sample.

Implications for Prevention Science

The field of prevention has moved into an era of Type II translational research in order to understand the gap between implementation quality achieved under real-world conditions as opposed to in the context of efficacy trials (Rohrbach et al., 2006). The current study advances this line of research by revealing potentially important associations between teacher and school contextual variables and implementation of PAX GBG that are consistent with prior research. In

addition to replicating these findings, an area for future research would be to explicitly examine whether efficacious interventions can be enhanced with readiness interventions to determine whether teachers varying in experience and perspectives can implement interventions equally well (Greenberg, Domitrovich, Graczyk, & Zins, 2005).

Acknowledgements

This research was supported in part by grants from the Institute of Education [R305A130060 and R305A080326] and the National Institute of Mental Health [P30 MH08643]. The authors would like to thank the administration and the teachers of the Baltimore City School District who are devoted educators of children. Dr. Domitrovich is an author on the PATHS Curriculum and has a royalty agreement with Channing-Bete, Inc. This has been reviewed and managed by Penn State's Individual Conflict of Interest Committee.

References

Baker, C. N., Kupersmidt, J. B., Voegler-Lee, M. E., Arnold, D. H., & Willoughby, M. T. (2010). Predicting teacher participation in a classroom-based, integrated preventive intervention for preschoolers. *Early Childhood Research Quarterly, 25,* 270-283.

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.

Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the pieces together: an integrated model of program implementation. *Prevention Science*, *12,* 23-33.

Beets, M. W., Flay, B. R., Vuchinich, S., Acock, A. C., Li, K.-K., & Allred, C. (2008). School climate and teachers' beliefs and attitudes associated with implementation of the positive action program: a diffusion of innovations model. *Prevention Science*, *9,* 264–75. doi:10.1007/s11121-008-0100-2.

Bevans, K. B., Bradshaw, C.P., Miech, R., & Leaf, P.J. (2007). Staff- and school-level predictors of school organizational health: A multilevel analysis. *Journal of School Health, 77*, 294-302. doi:10.1111/j.1746-1561.2007.00210.x

Bradshaw, C.P., Sawyer, A.L., & O'Brennan, L.M. (2009). A social disorganization perspective on bullying-related attitudes and behaviors: The influence of school context. *American Journal of Community Psychology, 43*(3-4), 204-220. doi:10.1007/s10464-009-9240-1

Bryk, A.S. & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. New York: Russell Sage.

Cameron, A. C. & Trivedi, P. K. (1998). *Regression Analysis of Count Data*. New York: Cambridge Press.

Chen, H.T. (1998). Theory-driven evaluations. *Advances in Educational Productivity, 7,* 15-34.

Choi, J. N. (2003). How does context influence individual behavior? Multilevel assessment of the implementation of social innovations. *Prevention and Treatment, 6*.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Belmont, CA: Wadsworth Group.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18,* 23-45.

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.

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

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. & Ialongo, N. (2008a). *The Teacher Perceptions of the Intervention Attributes (TPIA) scale.* Unpublished technical report. Johns Hopkins University.

Domitrovich, C. & Ialongo, N. (2008b). *The Perception of Administrative Support Scale (PASS).* Unpublished technical report, Johns Hopkins University.

Domitrovich, C., Poduska, J., & Bradshaw, C.P. (2008). *The Teacher-Coach Alliance Scale.* Unpublished technical report. Penn State University.

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*, 327-350.

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

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6*, 151- 175.

Gottfredson, D.C. & Gottfredson, G.D. (2002). Quality of school-based prevention programs: Results from a national survey. *Journal of Research on Crime and Delinquency, 39*, 3-35.

Gregory, A., Henry, D. B., & Schoeny, M. E. (2007). School climate and implementation of a preventive intervention. *American Journal of Community Psychology*, *40*, 250–60. doi:10.1007/s10464-007-9142-z

Greenberg, M. T., Domitrovich, C., Graczyk, P. A., & Zins, J. E. (2005). The study of implementation in school-based preventive interventions: Theory, research and practice. *Promotion of Mental Health and Prevention of Mental Behavioral Disorders, 3*, 1-62.

Greenberg, M.T., Kusché, C.A., & Conduct Problems Prevention Research Group (2011). Grade level PATHS (Grades 3-5). South Deerfield, MA: Channing-Bete Co.

Greenberg, M. & Kusché, C. (2006). Building social and emotional competence: The PATHS Curriculum. In S. Jimerson and M. J. Furlong (Eds.), Handbook of School Violence and Safety (pp. 395-412). New York: Lawrence Erlbaum Associates.

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

Hoy, W. K., & Feldman, J. (1987). Organizational health: The concept and its measure. *Journal of Research and Development in Education, 20,* 30-38.

Hoy, W. K., & Tarter, C. J. (1997). *The road to open and healthy schools: A handbook for Change, Elementary Edition.* Thousand Oaks, CA: Corwin Press.

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. 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. doi:10.1023/A:1021786811186.

Kellam, S.G., Poduska, J.M., Ialongo, N.S., Wang, W., Toyinbo, P., Petras, H., Ford, C., Windham, A., Wilcox, H.C. (2007). Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. *Drug and Alcohol Dependence, 95,* 5-28.

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

Land, K.C., McCall, P.L., & Nagin, D. (1996). A comparison of Poisson, negative binomial, and semiparametric mixed Poisson regression models: With empirical applications to criminal careers data. *Sociological Methods and Research, 24,* 387-442. doi: 10.1177/0049124196024004001

Luke, D. (2004). *Multilevel modeling*. Thousand Oaks, CA: Sage.

Main, S., & Hammond, L. (2008). Best practice or most practiced? Pre-service teachers' beliefs about effective behavior management strategies and reported self-efficacy. *Australian Journal of Teacher Education, 33,* 28-39.

Maslach, C., Jackson, S.E., & Leiter, M.P. (1997). Maslach burnout inventory: Third edition. In: Wood RJ, editor. *Evaluating stress: A book of resources*. Lanham, MD US: Scarecrow Education, p. 191-218.

Pas, E.T., Bradshaw, C.P., & Cash, A. (2014). Coaching classroom-based preventive interventions. In M. Weist, N. Lever, C.Bradshaw, & J. Owens (Eds.). *Handbook of School Mental Health, Second Edition* (pp.255-268). New York: Springer.

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,* 225–37. doi:10.1007/s11121-006-0029-2

Ransford, C. R., Greenberg, M. T., Domitrovich, C. E., Small, M., & Jacobson, L. (2009). The role of teachers' psychological experiences and perceptions of curriculum supports on the implementation of a social and emotional learning curriculum, *School Psychology Review, 38,* 510–532.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R., & du Toit, M. (2011). *HLM Statistical Software: Version 7*. Lincolnwood, IL: Scientific Software International, Inc.

Ringwalt, C. L., Ennett, S., Johnson, R., Rorhbach, L. A., Simons-Rudolph, A., Vincus, A., et al. (2003). Factors associated with fidelity to substance use prevention curriculum guides in the nation's middle schools. *Health Education and Behavior, 30,* 375–391.

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.

Rohrbach, L. A., Grana, R., Sussman, S., & Valente, T. W. (2006). Type II translation: Transporting prevention interventions from research to real-world settings. *Evaluation & the Health Professions, 29*, 302-333.

Schaffer, K., Rouiller, S., Embry, D., & Ialongo, N. (2006). *The PAX Good Behavior Game Implementation Rubric*. Unpublished technical report. Johns Hopkins University.

Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L.,…, & Saul, J. (2008). Bridging the gap between prevention research and practice: The Interactive Systems Framework for Dissemination and Implementation. *American Journal of Community Psychology, 41*, 171-181.

Wehby, J. H., Maggin, D. M., Moore Partin, T. C., & Robertson, R. (2011). The impact of working alliance, social validity, and teacher burnout on implementation fidelity of the Good Behavior Game. *School Mental Health, 4, 22-33.* DOI 10.1007/s12310-011-9067-4

Table 1.
Descriptive information on teacher participants and schools

	Percent	Mean (SD)		Range
Teacher-level variables (*n* = 199)				
Teacher is:				
Female	89			
In grades 3-5 classroom	43			
Young	45			
Has graduate degree	56			
Teacher reports of:				
Efficacy for behavioral management		3.87	0.6	2.21-5.00
Emotional exhaustion		3.39	1.38	1.00-6.56
Organization health (OHI)		2.91	0.47	1.78-4.00
Openness to innovation		3.29	0.51	2.00-4.00
Motivation to implement		3.91	0.98	0-5.00
Administrator support		3.07	1.14	0.50-5.00
Relationship with coach		4.45	0.66	1.00-5.00
Teacher feels GBG games:				
Fits with schedule		4.74	1.04	1.00-6.00
Fits with style		4.02	1.01	0-6.00
School-level variables (*n* = 18)				
Treatment status	50			
School size		357.78	(163.62)	205-941
FARMs rate		84.98	(7.66)	60.83-93.80
Percent of mobility		42.72	(8.88)	25.60-58.20
Percent of African American students		87.58	(21.02)	30.41-100
Percent of students suspended		7.70	(5.26)	0.30-18.22
Average OHI		2.94	(0.24)	2.57-3.44

Note. Young is coded (1= 20-30 years old, 0 = 31+); school size = the number of students enrolled; FARMs = free and reduced meals.

Table 2.
Hierarchical Linear Model Results for the Number of Games and Minutes Played Outcomes

	Games Played			**Minutes Played**		
	β	ERR	*p*	β	ERR	*p*
Intercept estimate (γ_{00})	6.12	453.78	<0.01*	8.81	6718.09	<0.01*
Individual Level						
Teacher is:						
Female (γ_{10})	-0.19	0.83	0.27	-0.09	0.92	0.67
In grades 3-5 class (γ_{20})	-0.15	0.86	0.11	0.08	1.08	0.52
Young (γ_{30})	-0.16	0.85	0.20	-0.15	0.86	0.36
Has graduate degree (γ_{40})	-0.07	0.94	0.51	-0.17	0.85	0.25
Teacher reports of:						
Efficacy for behavioral management (γ_{50})	-0.05	0.95	0.58	0.08	1.09	0.48
Emotional Exhaustion (γ_{60})	-0.10	0.91	0.02*	-0.08	0.93	0.16
Organization health- OHI (γ_{70})	-0.08	0.92	0.66	0.03	1.03	0.92
Openness to Innovation (γ_{80})	0.06	1.06	0.67	-0.04	0.96	0.82
Motivation to implement (γ_{90})	0.09	1.09	0.20	0.17	1.18	0.07
Administrator support (γ_{100})	-0.05	0.95	0.26	-0.05	0.95	0.46
Relationship with coach (γ_{110})	0.10	1.11	0.31	0.05	1.05	0.76
Teacher feels intervention:						
Fits with schedule (γ_{120})	0.00	1.00	0.94	0.05	1.05	0.59
Fits with style (γ_{130})	0.17	1.18	0.02*	0.16	1.17	0.08
School Level						
Treatment status (γ_{01})	-0.02	0.98	0.88	0.08	1.08	0.59
School size (γ_{02})	0.00	1.00	0.27	0.00	1.00	0.41
FARMs rate (γ_{03})	0.00	1.00	0.86	0.01	1.01	0.51
Average OHI (γ_{04})	-0.15	0.86	0.62	-0.71	0.49	0.07

Note. ERR = event rate ratio (referred to in other programs as incidence rate ratio). Young is coded (1= 20-30 years old, 0 = 31+); school size = the number of students enrolled; FARMs = free and reduced meals.
* $p < .05$.

Table 3.
Hierarchical Linear Model Results for the Game Observations Outcome

	Game Observations			
	Intercept (ψ_0)		Time Slope (ψ_1)	
	β	*p*	β	*p*
Intercept estimate (β_{000})	3.22	<0.01*		
Time estimate (β_{100})			0.11	0.15
Individual Level				
Teacher is:				
Female (β_{010})	-0.13	0.38	0.03	0.66
In grades 3-5 class (β_{020})	-0.03	0.79	0.01	0.77
Young (β_{030})	0.22	0.04*	-0.07	0.07
Has graduate degree (β_{040})	-0.14	0.17	0.00	0.91
Teacher reports of:				
Efficacy for behavioral management (β_{050})	0.04	0.57	0.00	0.93
Emotional Exhaustion (β_{060})	-0.02	0.69	0.01	0.66
Organization health - OHI (β_{070})	-0.14	0.37	0.01	0.84
Openness to Innovation (β_{080})	-0.04	0.77	0.01	0.82
Motivation to implement (β_{090})	0.06	0.34	-0.01	0.62
Administrator support (β_{0100})	-0.05	0.25	0.00	0.85
Relationship with coach (β_{0110})	0.08	0.36	0.01	0.82
Teacher feels intervention:				
Fits with schedule (β_{0120})	-0.04	0.57	0.02	0.56
Fits with style (β_{0130})	0.03	0.68	0.00	0.91
School Level				
Treatment status (β_{001})	0.07	0.60	-0.07	0.16
School size (β_{002})	0.00	0.21	0.00	0.53
FARMs rate (β_{003})	0.00	0.78	0.00	0.62
Average OHI (β_{004})	-0.18	0.58	0.11	0.34

Note. Young is coded (1= 20-30 years old, 0 = 31+); school size = the number of students enrolled; FARMs = free and reduced meals. Unconditional intra-class correlation (ICC) = 0.18, Final ICC = 0.10 (45.9% reduction in ICC). $^{*}p < .05$.