ORIGINAL PAPER

Identifying Patterns of Coaching to Support the Implementation of the Good Behavior Game: The Role of Teacher Characteristics

Elise T. Pas · Catherine P. Bradshaw ·
Kimberly D. Becker · Celene Domitrovich ·
Juliette Berg · Rashelle Musci · Nicholas S. Ialongo

Abstract   There is growing interest in coaching to support teacher implementation of evidence-based interventions; yet, there is limited research examining the tailoring of coaching support to teachers' needs. This paper examined coaching dosage across one school year, and the relationship between coaching contacts and teacher baseline and end-of-year data. Data came from a randomized controlled trial including 210 teachers in 18 schools implementing the Good Behavior Game (GBG), either as a stand alone or when integrated with a social–emotional learning curriculum. The overarching goal was to determine whether coaches provided varying levels of teacher contacts and how this support related to condition assignment, implementation, and teachers' beliefs and perceptions data. Growth mixture modeling (GMM) was used to examine the frequency of teacher contacts across the school year. GMM indicated three distinct patterns: about 58 % of teachers received a moderate number of contacts; 27 % received a consistently low number of contacts; and 15 % received high and increasing support. Teachers who received a high degree of support were more often implementing the integrated GBG and reported more negative beliefs and perceptions at the start of the school year than those in the low contact class. Teachers in the low contact class implemented the least number of games and minutes of GBG, but reported better perceptions of organizational health and burnout, at the end of the year. Coaching dosage was unrelated to observer ratings of implementation quality.

Introduction

Schools serve as an ideal setting to provide evidence-based interventions (EBIs) given their access to all children, particularly in disadvantaged neighborhoods, and the high prevalence of student behavioral and mental health challenges (Hoagwood et al., 2007); however, limited adoption and poor implementation of EBIs in schools are problematic (Dusenbury, Branningan, Falco, & Hansen, 2003; Hicks, Shahidullah, Carlson, & Palejwala, 2014; Ringwalt et al., 2003). In recent years, federal agencies, researchers, and policymakers have shown increasing interest in implementation science, which encompasses studies examining real-world implementation of EBIs (Spoth et al., 2013). Although there is interest in coaching models as a means for promoting teacher development and implementation (Becker, Bradshaw, Domitrovich, & Ialongo, 2013a), additional research is needed to better understand the frequency and duration of supports needed to optimize teacher implementation. The current paper examined specific patterns of coaching dosage provided to teachers implementing an EBI, called the PAX Good Behavior Game (GBG; Barrish, Saunders, & Wolf, 1969; Embry,

E. T. Pas (✉) · C. P. Bradshaw · R. Musci · N. S. Ialongo
Johns Hopkins University, Bloomberg School of Public Health,
Baltimore, MD, USA
e-mail: epas@jhu.edu

C. P. Bradshaw · J. Berg
University of Virginia, Charlottesville, VA, USA

K. D. Becker
University of Maryland, School of Medicine, Baltimore, MD,
USA

C. Domitrovich
Pennsylvania State University, University Park, PA, USA

Staatemeier, Richardson, Lauger, & Mitich, 2003). PAX GBG is designed to promote student self-regulation and prevent student disruptive behavior. In this study, we also tested a condition where the GBG was combined with a social–emotional learning program.

The overall purpose of the current paper was to determine whether differentiated coaching dosage aimed at improving teacher implementation was associated with teachers' implementation of the PAX GBG as well as beliefs and perceptions about themselves and the school environment (e.g., burnout, efficacy, organizational health). We were particularly interested in variation in implementation quality and dosage of the GBG when implemented as a stand-alone intervention in contrast to the integrated condition, where it was combined with the promoting alternative thinking strategies (PATHS) social–emotional learning curriculum (Greenberg, Kusché, & CPPRG, 2011; Kusché, Greenberg, & CPPRG, 2011). The goal of the coaching was to support teachers in their implementation of these two EBIs.

Need for Behavioral Interventions in Schools

Teachers' inability to effectively address behavior problems is among the leading reasons for teacher turnover and exiting of the profession (Ingersoll & Smith, 2003). In fact, teacher turnover is estimated to occur among approximately 10 % of public school teachers in their first year, and an additional 12 % leave after 2 years of teaching (Kaiser & Cross, 2011). This turnover disrupts continuity in the educational workforce and creates a need for ongoing training within schools to ensure that staff are prepared to consistently address student needs. These issues are compounded by the fact that many teachers lack classroom and behavioral management training and have a desire for additional support in this area (Baker, 2005; Reinke, Stormont, Herman, Puri, & Goel, 2011; Siebert, 2005). This cycle of lack of preparation, an inability to address student needs, and turnover creates ongoing challenges for schools trying to implement complex systems of support targeting student mental health, resulting in difficulties addressing the needs of the most "at-risk" students (i.e., those with more intensive needs and thus interventions).

Concerns with School-Based Implementation

Although randomized trials suggest that there are a number of effective preventive interventions available for use by schools (for reviews, see Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Hoagwood & Burns, 2005; O'Connell, Boat, & Warner, 2009; Wilson & Lipsey, 2007), implementation in real-world settings warrants concern at both the adoption and implementation levels

(Dusenbury et al., 2003; Gottfredson, Jones, & Gore, 2002; Hicks et al., 2014; Ringwalt et al., 2003). In terms of adoption, research suggests that fewer than 10 % of the programs implemented in schools to prevent drug use and crime are research-based and less than half of all schools implement at least one research-based program (Ringwalt et al., 2011; U.S. Department of Education, 2012). Even when EBIs are adopted, fewer than half of programs are implemented with minimal fidelity (Ringwalt et al., 2011; U.S. Department of Education, 2012).

Enhancing Implementation Through Coaching

Theoretical models of implementation indicate that ongoing support is needed in schools to ensure implementation and that these supports should be considered as essential components of the implementation process. For example, Wandersman et al. (2008) highlight the need for a system to deliver interventions, to facilitate the occurrence of implementation, and to prepare interventions to be received and utilized by users. This issue is further emphasized by Domitrovich et al. (2008), who presented macro-, school-, and individual-level factors impacting implementation and the subsequent supports needed to ensure consistent implementation quality. Like Wandersman et al. (2008), Domitrovich et al. highlighted that the intervention and support systems are two layers to be considered in the implementation model. Coaching is a means to support the implementation of EBIs, which meets these theoretically indicated systems and supports (Becker, Darney, Domitrovich, Keperling, & Ialongo, 2013b).

There is increasing interest in coaching models as a means for developing teacher skills as well as to provide support in the implementation of a specific program (e.g., reading coaches) or target areas (e.g., behavioral coaches; Pas, Bradshaw, & Cash, 2014). Yet there remain a number of gaps in the research on coaching. For example, there is wide variability in the duties performed by coaches (Becker et al., 2013a), as well as the structure by which coaching is provided (Pas & Newman, 2013). There is also a shortage of rigorous research isolating the effects of coaching supports; and when found to be effective, it is not clear what coaching dosage is necessary to change teachers' behavior and, in turn, improve outcomes for students. Further, there are gaps in the literature regarding the effectiveness of coaching generally, particularly in the area of coaching to promote EBI implementation.

One recent study examined coaching as a means for promoting the implementation of GBG and showed that the alliance formed between the coach and teacher was the strongest predictor of implementation, controlling for a number of teacher variables (Wehby, Maggin, Moore Partin, & Robertson, 2012). In addition, coaching was

found to buffer the association between teacher stress and implementation. Another study on the PAX GBG program drew upon data from the current data set and focused on coaches' efforts to tailor their use of specific coaching practices (e.g., modeling, feedback, delivery) in order to improve the quality of teachers' implementation (Becker et al., 2013a). The authors found that coaches strategically varied their coaching practices based on implementation quality and the rollout of the PAX GBG program over the course of the school year. Although these prior studies provided important information regarding factors that influence coaching, further research is needed regarding the dosage of coaching and how it is tailored based on teacher characteristics and implementation in order to promote EBI fidelity.

Current Study

The current study examined the patterns of coaching dosage provided to teachers involved in a randomized controlled trial (RCT) across one school year with the goal of exploring the extent to which coaching dosage related to teacher baseline and end-of-year data. As noted earlier, the RCT tested the PAX GBG (Embry et al., 2003) as implemented alone and integrated with the PATHS curriculum (Greenberg et al., 2011; Kusché et al., 2011). Both PAX GBG and PATHS are evidence-based preventive interventions implemented in schools. The original GBG was developed by Barrish et al. (1969) as a classroom management strategy; Embry et al. (2003) augmented the original GBG, referred to as PAX GBG, by incorporating research-based additions of teacher and student verbal and visual cues to promote attentive and prosocial behaviors outside of the formal games (i.e., with the goal of improving generalization). Prior studies testing the original GBG have demonstrated positive academic, behavioral, and substance use outcomes (e.g., Ialongo et al., 1999; Kellam et al., 2008). The PATHS program includes developmentally appropriate lessons and activities providing direct instruction and practice opportunities to develop students' social–emotional skills. PATHS program has been experimentally tested and demonstrated reductions in off-task, aggressive, and disruptive student behaviors through the improvement of prosocial cognitions and socially competent behaviors (e.g., CPPRG, 2010; Greenberg, Kusché, Cook, & Quamma, 1995). In the current study, our focus was on the implementation quality and dosage of GBG, regardless of condition (PAX GBG alone or PAX GBG integrated with PATHS), because it was the common EBI across both conditions.

Building on prior work with this sample, which focused on the types of coaching supports provided to optimize implementation quality (Becker et al., 2013a, 2013b), the current study focused on the association between varying dosage of coaching support and teachers' implementation of the program. Both experimental conditions were included in the analysis in order to examine whether coaching supports were provided differently based on whether teachers were in the PAX GBG only or integrated condition. As such, we hypothesized that the integrated condition, which included a greater number of implementation components and may have been more complex and challenging for teachers to master, would require more coaching supports and be associated with less GBG implementation quality and dosage, relative to the stand-alone PAX GBG condition.

We were also interested in the extent to which the coaching dosage over the year varied as a function of teacher characteristics and their beliefs and perceptions of themselves and the school. Although the coaching model followed a coaching manual [see description of coaching model in the methods section and Becker et al. (2013a, 2013b)], which included universal supports to all teachers, tailoring and thus variation in coaching supports to meet the teachers' needs were expected. Therefore, we hypothesized that there would be some variation in the level of support provided to teachers (e.g., low, moderate, and intensive levels of support) and that these patterns or trajectories of coaching support over time would be functionally associated with baseline teacher characteristics as well as teacher-reported end-of-year data. Specifically, we hypothesized that the level of support would be associated with level of teacher need, as measured by implementation quality at the beginning of the year as well as baseline teachers' experience of burnout, efficacy, and perceptions of the school environment. Furthermore, we hypothesized that a greater level of support would be associated with improved implementation and teacher beliefs and perceptions at the end of the school year.

Method

Design Overview

Data for this study were drawn from a preventive intervention RCT where elementary schools were randomized to one of three conditions: the integrated (PATHS/GBG; nine schools) condition where teachers implemented PAX GBG with the PATHS program (Greenberg et al., 2011; Kusché et al., 2011), PAX GBG only (nine schools), and a control condition (nine schools) where teachers conducted their usual practice. The current study included the 18 intervention schools where teachers received coaching to support their implementation of PAX GBG only and the integrated model. The inclusion of both intervention

conditions allows for the exploration of whether teachers implementing the integrated intervention received more intensive coaching.

The study was conducted in a large urban, east coast public school district (see school demographics in Table 1). The majority of students in this sample was African American ($M = 88$ %) and received free and reduced meals ($M = 85$ % across schools, range = 70.2–95.5 %). Schools were recruited and principals provided their agreement to participate, be randomized, and potentially receive 1 year of training and coaching in the interventions. Schools and teachers were enrolled during three consecutive school years (i.e., cohorts) for their 1-year participation (i.e., each year, six schools participated in the study, two of which were assigned to each condition). Teacher participation in the intervention and data collection was voluntary and consent was provided. The IRB at the principal investigators' institution approved this study.

Participants

The original study sample included 222 K-5 intervention teachers across the 18 intervention schools. Schools and

therefore teachers were enrolled in three cohorts (i.e., for 1 year each, in three consecutive years) of approximately equal sizes (33 % in cohort 1, 36.5 % in cohort 2, and 30.2 % in cohort 3). Approximately 55 % of the teachers ($n = 121$) were in schools assigned to receive PAX GBG only and 45 % were in the PATHS/GBG condition ($n = 101$). The vast majority of teachers were women (i.e., 87 %) 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 teacher sample, as well as average scores on the variables measured in this study. Among these 222 teachers, 210 teachers both received coaching supports (i.e., 216) and consented to provide survey data (i.e., six teachers were coached but did not consent to provide survey data). Therefore, the final sample for the current analysis was 210 teachers.

Interventions

All intervention teachers were trained to implement PAX GBG. The GBG uses a team-based, game-like context to promote self-regulation and reduce aggressive, disruptive, and off-task behavior and thus facilitate academic

Table 1 Descriptive information on teacher participants and schools

	Percent	Mean (SD)	Range
Teacher-level variables ($n = 222$)			
Teacher			
Is a woman	87		
Teaches grades 3–5	44		
Is younger aged	44		
Has graduate degree	57		
Teacher baseline reports of:			
Behavior management efficacy		3.87 (0.62)	1.86–5.00
Social–emotional learning efficacy		3.64 (0.65)	1.63–5.00
Emotional exhaustion		3.38 (1.40)	1.00–6.89
Depersonalization		2.21 (1.28)	1.00–6.33
Personal accomplishment		5.89 (0.85)	3.00–7.00
Organization Health Inventory		2.92 (0.49)	1.58–4.00
Teacher implementation:			
Wave 1 Implementation Rubric score		3.11 (0.71)	0.00–4.00
Wave 4 Implementation Rubric score		3.33 (0.58)	0.71–4.00
Number of PAX GBG games played		151.30 (99.52)	1–433
Minutes of PAX GBG played		1,484.03 (1,373.65)	5–7,388
School-level variables ($n = 18$)			
Treatment status	50		
School size		357.78 (163.62)	205–941
Free and reduced meals rate		84.98 (7.66)	60.83–93.80
Percent of mobility		42.72 (8.88)	25.60–58.20
% African American students		87.58 (21.02)	30.41–100
Percent of students suspended		7.70 (5.26)	0.30–18.22
Mean score: Organization Health Inventory		2.94 (0.24)	2.57–3.44

Younger aged was coded (1 = 20–30 years old, 0 = 31+); school size = the number of students enrolled

instruction. The integrated condition used in nine of the schools had teachers implement both PATHS (Greenberg et al., 2011; Kusché et al., 2011) and PAX GBG (see Domitrovich et al., 2010 for description). Teachers in the PAX GBG only condition received 1.5 days of training (i.e., 1 full day and then a half day booster), whereas the integrated condition received 3.5 days of training (i.e., 2 days of PATHS and 1.5 days of PAX GBG). As noted above, implementation of GBG was the common component across the two conditions and thus was examined in the current study.

Overview of the Coaching Model

After participating in the initial group training, all participating teachers received face-to-face coaching for the entire school year. Coaching had manualized components, but the intensity was tailored to individual teacher needs. Detailed coaching and implementation data were collected during an intervention period of approximately 31 weeks, during which coaches were expected to meet with each teacher approximately once a week. Coaches followed a two-phased coaching model (for additional details, see Becker et al., 2013; Becker et al., 2013). The first phase involved the manualized universal coaching phase lasting approximately 4–6 weeks after the intervention workshop trainings, during which coaches used the same coaching strategies with all teachers (e.g., check-ins, modeling, needs assessments, and technical assistance/performance feedback). At the end of this 6-week period, coaches accompanied members of the research team while they conducted the first of four independent observations of teachers' program delivery and completed an implementation rubric rating (see description below).

The second phase was a dynamic one whereby standard components were implemented, but the intensity and frequency of their use were tailored based on ongoing observations including the rubrics, implementation dosage via teacher report of frequency of games played, and teacher requests for help with a specific problem(s). Specifically, coaches were expected to continue to collect data from teachers on a weekly basis about the number and duration of games played (i.e., dosage). Coaches developed individualized plans for each teacher regarding additional tailored contacts using these weekly dosage data as well as informal structured observations and formal implementation rubrics that were administered at the end of four, roughly quarterly, waves (i.e., fall, winter, early spring, late spring). Coaches followed the manual by having some contact (e.g., in-person, e-mail, phone) with all teachers on a weekly basis; the manual also promoted consistency for when significant benchmarking activities (e.g., introductions, needs assessments, and formal observation of

implementation) occurred [see Becker et al. (2013a, 2013b) for additional details on the coaching model]. The frequency, intensity, and nature of the activities based on teachers' level of skill and use of the GBG varied. In practice, the number of coach contacts with teachers also varied based on teacher receptivity to coaching and idiosyncratic factors, such as severe weather that disrupted the school calendar and teacher absence (e.g., leave time and illness).

Coaches were hired, trained, and supervised by the research team. All coaches were former teachers and had experience implementing the PAX GBG intervention. Though they were external providers to the schools (i.e., were hired by the research team), the coaches functioned as other support staff (e.g., school psychologists) in that they traveled freely across classrooms and schools. They had school assignments (averaging 2–3 schools per year) and regular access to teachers. They scheduled coaching sessions with the teachers based on mutual availability.

Measures

Coaching Contacts

After each in-person visit, the coaches recorded details about the services provided to the school (i.e., total time and activity type). Only substantive in-person contacts of at least 5 min or longer were recorded. The number of contacts for each of the four data collection waves was calculated for each teacher and used to assess coaching dosage. These data were used in the modeling of coaching dosage trajectories.

Quality of PAX GBG Implementation

Rubric ratings of teachers' PAX GBG game quality were completed by coaches and other research staff during four waves throughout the academic year. During this observation, teachers were asked to conduct a 5- to 10-min game so that the observer could determine whether elements were properly executed. The Game Observation Scale of the PAX GBG rubric (Schaffer, Rouiller, Embry, & Ialongo, 2006) included seven items assessing teacher preparation for and execution of the game ($\alpha = .93$). This included: (1) preparing students, (2) the activity during which the game is conducted, (3) timer usage, (4) team structure, (5) teacher response to behavior, (6) game review at the end, and (7) the prize given. Ratings were made on a scale of 0–4, with higher scores indicating better implementation. The assessment team was comprised of coaches and research staff who were randomly assigned to complete observations, except in the first cohort during the first wave. Inter-rater reliability was assessed using pairs of a

coach and staff member for the first 15 % of teachers at each data collection wave. Once a reliability of .80 or higher for each item was achieved, remaining observations were conducted by only one person. Independent observer data for the implementation rubric were used for these analyses, even when two observers were present.

Dosage of PAX GBG

Each week, teachers completed and submitted a log of the number of games played and the duration of each game played. These data were summed across the school year and yielded two variables: total number of games implemented and total number of minutes implementing PAX GBG that were used summatively across the entire year.

Teacher Demographics

Teachers provided information regarding their demographic data (i.e., gender, age, education, years teaching, degree attainment) on a teacher information form at the start of the study.

Beliefs and Perceptions Data

Teachers provided baseline and end-of-year ratings of their beliefs and perceptions across a variety of domains including efficacy, burnout, and school environmental factors. Of specific interest in the current study were two scales of efficacy: the Behavior Management Self-Efficacy Scale (Main & Hammond, 2008), which included 14 items specific to promoting classroom behavior management (e.g., "I am able to use a variety of behavior management techniques"; $\alpha = .94$) and the Social–Emotional Learning Self-Efficacy Scale (Domitrovich & Poduska, 2008), which included eight items regarding efficacy in promoting students' social–emotional development (e.g., "I am able to use a variety of techniques to teach children positive social skills"; $\alpha = .93$). Item responses for both efficacy scales were provided on a 5-point Likert-type scale ranging from "not at all" to "a lot."

We also administered the three scales of the teacher report version of the Maslach Burnout Inventory (MBI; Maslach & Jackson, 1981; Maslach, Jackson, & Leiter, 1997), including emotional exhaustion (nine items, e.g., "I feel used up at the end of the workday", $\alpha = .92$), depersonalization (three items, e.g., "I worry that this job was hardening me emotionally", $\alpha = .64$), and personal accomplishment (eight items, e.g., "I feel I'm positively influencing my students' lives through my work", $\alpha = .82$). Responses were rated on a 7-point scale from "never" to "every day."

Finally, teachers completed the Organizational Health Inventory (OHI; Hoy & Feldman, 1987), which included 31 items assessing the organizational health of the school across four domains: teacher affiliation (nine items), academic emphasis (five items), collegial leadership (ten items), and resource influence (seven items; Hoy & Tarter, 1997). Responses were provided on a 4-point Likert-type scale ranging from "rarely occurs" to "very frequently occurs." A total score for the OHI ($\alpha = .93$) was calculated by averaging the responses across all items for each teacher.

Analyses

Growth mixture modeling (GMM) was conducted in M*plus* 7.1 (Muthén & Muthén, 1998–2013) to assess the growth trajectories for number of coach contacts with teachers across four data collection waves within the school year. The four waves were not equally distributed across the school year due to the school calendar (e.g., winter break, spring break, testing); for example, the two spring waves were in closer proximity than the first fall and second winter waves. Nevertheless, the four time frames were used to track the rollout and implementation of the program across the school year and map onto the collection of implementation quality rubrics.

The GMMs were built iteratively, such that one growth class was added at a time and the total number of classes was determined using three fit indices and two statistical tests: Akaike information criteria (AIC), Bayesian information criterion (BIC; Schwartz, 1978), sample size adjusted BIC, Lo–Mendell–Rubin likelihood ratio test (LMR; Lo, Mendell, & Rubin, 2001), and Vuong-LMR likelihood ratio test (Muthén & Muthén, 1997–2012). A decreasing AIC, BIC, and adjusted BIC and statistically significant LMR and Vuong-LMR indicated improved fit of the addition of a growth class. Further, entropy scores were examined with a focus on attaining scores closest to 1.00 and latent class probabilities >.70 (Nagin, 2005; Ramaswamy, DeSarbo, Reibstein, & Robinson, 1993). Finally, an additional class was added only if the solution resulted in classes of meaningful size and conceptual and theoretical relevance (Muthén, 2004; Nylund, Asparouhov, & Muthén, 2007).

Once the GMM was finalized, predictors of class membership (i.e., intervention condition, demographics, baseline rubric score, and baseline beliefs and perceptions) were added to the model. Specifically, the latent GMM classes were regressed on all of these targeted variables using a multinomial logistic regression framework. Finally, in order to be included in the GMM as an outcome, the end-of-year data were converted into binary latent class variables; these variables were created to indicate high versus low levels of each outcome and were regressed on the growth classes, while controlling for the predictors.

Specifically, we examined end-of-year dosage (i.e., number of games and minutes), quality of implementation (i.e., rubric score), and beliefs and perceptions (i.e., efficacy, burnout, and organizational health). Each end-of-year variable was dichotomized such that $0 =$ below the 66th percentile and $1 =$ at or above the 66th percentile, to represent a benchmark for high (i.e., score of 1) versus low (i.e., score of 0) implementation quality, dosage, and beliefs and perceptions. Specifically, a 1 code corresponded to the highest dosage, highest quality, highest efficacy, highest burnout (i.e., the one undesirable outcome in this coding category), and highest/best perception of the school environment. These values were chosen in the absence of an otherwise-established cutpoint, while also ensuring that there were enough teachers in each group to be compared statistically. Each end-of-year variable was modeled separately resulting in nine final models; the GMM and baseline covariates portions of the model were identical for each model and only the outcome variable changed.

Missing Data

The rates of missingness on individual beliefs and perceptions and implementation fidelity (i.e., dosage and quality) variables ranged from 0 to 22 %. Therefore, to ensure a complete dataset where no cases were dropped, all missing data were imputed using a multivariate imputation by chained equation (MICE) method of multiple multivariate imputation in STATA (Azur, Stuart, Frangakis, & 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; for which we had complete data). School-level predictors (which had complete data) such as school size (i.e., enrollment), free and reduced meals, and mobility

were also included to inform the imputation. All teacher-level beliefs and perceptions variables were imputed using MICE.

Results

Coaching Trajectories

A series of GMMs with up to five latent growth classes was fit using the total number of coach contacts with teachers prior to each of the four data collection waves (see Table 2). The best fit for the GMM of coach contacts with teachers included three growth trajectories (LMR $p < .01$ for 3-class solution, entropy $= .98$; see Table 2 for fit statistics and Fig. 1 for a graphical depiction of the three-class model). Although the 4-class model indicated improved fit indices and significant tests of model fit, the 4-class solution resulted in a substantial drop in entropy, errors in the convergence of the log-likelihood values, and a fourth class that was comprised mainly of a second *high* class (i.e., taking the one *high* class seen in the 3-class model and creating two classes that were largely non-distinguishable); thus, it lacked added meaning.

The largest latent growth class in the 3-class solution, comprising 57.9 % of the sample, consisted of teachers receiving *moderate and steady* levels of contacts across the year with the most stability in waves 1–3 (i.e., approximately 8–10 contacts each wave) and a slight tapering of support at the end of the year to about five contacts prior to the fourth data collection wave. This class is referred to hereafter as *moderate*. The next largest growth class (i.e., 27.3 % of the sample) of teachers received *consistently low* coach contacts (referred to hereafter as *low*), starting with about three contacts in the first wave and ending with about one in the final wave. The final growth class received *high and increasing* levels of supports (referred to hereafter as *high*). Coaches visited these teachers (i.e., 14.7 % of the

Table 2 Fit statistics for growth mixture model for trajectories of coach contacts with teachers

	1-Class	2-Class	3-Class	4-Class	5-Class
Log likelihood	−3078.39	−2381.31	**−2196.47**	−2173.92	−2153.27
AIC	6162.78	4776.61	**4414.93**	4377.85	4344.55
BIC	6172.90	4800.24	**4452.06**	4428.48	4408.68
Adjusted BIC	6163.40	4778.06	**4417.21**	4380.94	4348.47
VLMR		0.00	**0.00**	0.00	0.19
LMR		0.00	**0.00**	0.00	0.20
Entropy		0.97	**0.98**	0.86	0.88

Bold indicates best fitting model

AIC Akaike information criteria, *BIC* Bayesian information criterion, a*djusted BIC* sample size adjusted BIC, *VLMR* Vuong Lo–Mendell–Rubin likelihood ratio test, *LMR* Lo–Mendell–Rubin likelihood ratio test
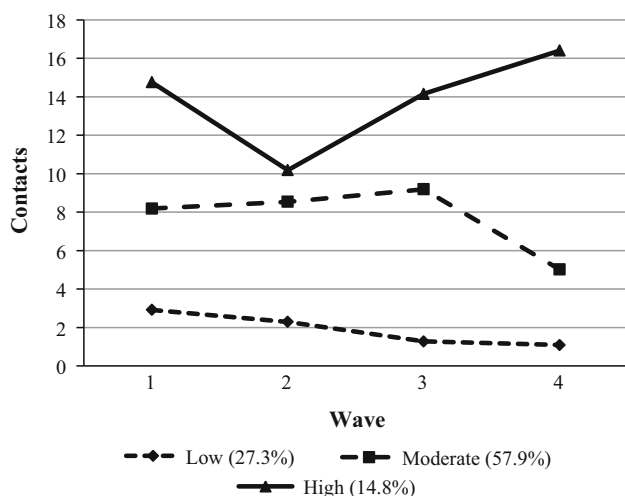
Fig. 1 Sample means for the growth classes of the 3-class solution. Values on the y-axis represent average number of coach contacts with the teachers. The x-axis includes the four waves across the school year

sample) about 15 times in the first wave, decreased to about 10 visits in the second wave, and then increased the number of contacts again in the final two waves.

Relationships Between Coaching Trajectories and Baseline Teacher Variables

Predictors of the growth trajectory of coaching dosage were next added to the model. Specifically, the treatment condition (i.e., PAX GBG vs. integrated) as well as other design variables (cohort/year in which the teacher participated in the study; the coach the teacher predominantly worked with), demographics, baseline rubric score, and teacher-reported beliefs and perceptions were included in the model. In comparing the *low* contact class to the *moderate* class, we found no significant differences on these predictors. A number of differences between the *low* and *high* classes emerged (see Table 3 for all results). As we hypothesized, teachers receiving a high number of contacts, as compared only to those receiving a low number of contacts, were less likely to be in the PAX GBG only condition than in the integrated condition ($\beta = -53.96$, $p < .01$). We also found that teachers in the *high* contact class were also more likely to have participated in years one ($\beta = 9.57$, $p < .01$) and three ($\beta = 363.21$, $p < .01$) of the study and were more likely to have worked with two of the three coaches specifically (βs = 284.34 and 268.79, ps $< .01$). With regard to teacher-specific variables, women were more likely to be in the *high* class ($\beta = 11.54$, $p = .03$), whereas teachers under the age of 30 ($\beta = -17.83$, $p < .01$), and those teaching in grades 3–5 ($\beta = -21.62$, $p = .01$) were less likely to be in the *high* class. The teachers in the *high* contact class also had higher

baseline ratings of depersonalization (i.e., more burnout; $\beta = 18.84$, $p = .01$) than the *low* class and lower ratings of efficacy for behavioral management ($\beta = -70.73$, $p = .03$). No other predictors were significantly associated with the coaching trajectories (see Table 3).

Outcomes Regressed on Coaching Trajectories

As stated earlier, all end-of-year variables had to be modeled separately. The reference group for all analyses was set to be the *low* contact class (see Table 4 for all findings). A number of variables measured at the end of the year, including PAX GBG game dosage and teacher beliefs and perceptions, were related to the coaching dosage received. Teachers in the *moderate* ($\beta = 0.96$, $p = .02$) and *high* contact ($\beta = 1.29$, $p = .02$) classes were more likely to implement a high number of games, as compared to the *low* class. Similarly, the *high* contact class was more likely than the *low* class to implement a high number of minutes of games ($\beta = 1.55$, $p < .01$). The comparison of the *moderate* contact class, as compared to the *low* class, approached significance, showing a trend of playing the game for more minutes ($\beta = 0.78$, $p < .10$). In other words, teachers in the *low* contact class demonstrated the highest likelihood of being low dosage implementers as compared to teachers in the other two classes. Ratings of quality using the rubric in the final wave were not significantly different for teachers across classes.

With regard to the teachers' self-reported data, teachers receiving a moderate number of contacts ($\beta = -0.89$, $p = .02$) were less likely to provide high ratings of the organizational health as compared to the *low* contact class. The *moderate* contact class was also more likely to report high levels of depersonalization ($\beta = 1.44$, $p < .01$) as compared to the *low* contact class. Importantly, these two differences (i.e., on organizational health and depersonalization) were not present between these classes at baseline. End-of-year ratings of behavioral management efficacy, emotional exhaustion, and personal accomplishment were not related to class membership.

Discussion

This study built on prior work which examined types of coaching supports in relation to implementation quality using this sample (Becker et al., 2013a) by examining coaching supports over time in relation to implementation dosage, quality, and teachers' beliefs and perceptions about themselves and the school. We were particularly interested in whether the receipt of differing levels of coaching was associated with teacher characteristics, teachers' beliefs and perceptions, and implementation at the start and end of

Table 3 Association between baseline variables and being in the moderate and high (vs. low) contact class

	Moderate			High		
	Logistic coefficient	SE	*p* value	Logistic coefficient	SE	*p* value
Coach 1±	10.26	8.60	0.23	284.34	43.43	0.00*
Coach 2±	−3.92	2.10	0.06	268.79	58.15	0.00*
Cohort 1±	−14.73	9.43	0.12	9.57	0.16	0.00*
Cohort 3±	−3.03	1.86	0.10	363.21	79.30	0.00*
PAX GBG only (vs. integrated)	1.28	2.25	0.57	−53.96	12.24	0.00*
Implementation Rubric score[†]	−0.75	1.04	0.47	20.33	15.03	0.18
Organizational Health Inventory[†]	0.65	1.01	0.52	−21.11	22.42	0.35
Behavior management efficacy[†]	4.34	2.99	0.15	−70.73	33.29	0.03
Social–emotional learning efficacy[†]	0.34	1.94	0.86	−2.40	8.39	0.78
Emotional exhaustion[†]	0.29	0.56	0.61	4.29	5.94	0.47
Personal accomplishment[†]	−1.51	1.45	0.30	−12.45	11.32	0.27
Depersonalization[†]	−0.85	0.62	0.17	18.84	7.14	0.01*
Woman	3.46	2.04	0.09	11.54	5.25	0.03*
Teaches grades 3–5 (vs., K-2)	−2.67	2.66	0.32	−21.62	7.80	0.01*
Young teacher (≤ 30)	−0.14	1.32	0.92	−17.83	1.41	0.00*
Graduate degree (vs., no advanced degree)	−3.72	3.18	0.24	0.47	0.75	0.53

The *consistently low* class serves as the reference group for these analyses

* *p* < .05

± Dummy coded variable

[†] Assessed at baseline. Estimates reported are unstandardized multinomial logistic coefficients

Table 4 Association between class membership and being above the 66th percentile for end-of-year outcome data

	Moderate	High and increasing
Implementation Rubric score	−0.040	0.031
Number of PAX GBG games	0.957*	1.293*
Number of minutes spent playing PAX GBG	0.779[†]	1.555*
Behavior management efficacy	−0.473	0.088
Social–emotional learning efficacy	−0.567	0.377
Emotional exhaustion	0.744	0.671
Depersonalization	1.442*	0.493
Personal accomplishment	0.327	−0.219
Organizational Health Inventory	−0.888*	−0.718

Each end-of-year outcome variable was modeled separately. The reference group, in all cases, is the *consistently low* contact class. Positive coefficients represent that the group is more likely to have scores above the 66th percentile (i.e., a desired outcome for all but the three burnout scales; emotional exhaustion, depersonalization, and personal accomplishment)

[†] *p* < .10; * *p* < .05

the school year. We also sought to examine whether the type of the intervention (i.e., PAX GBG only vs. integrated intervention) was associated with the level of coaching support provided, given the additional components of an integrated intervention. The findings generally indicated that teachers did in fact receive varying starting levels and trajectories across the school year of coaching supports. Specifically, over half of the teachers were in the *moderate* class, receiving regular though not an intensive amount of coach contacts during each wave. Just about 15 % of teachers received intensive, high levels of coaching support. About a quarter of the teachers were in a *low* contact class, receiving only a couple of coach visits during each wave, demonstrating that a rather sizable portion of the sample (27 %) received a somewhat 'lighter touch' than outlined in the coaching model (Becker et al., 2013b).

To better understand the extent to which the classes may have been tailored to map onto the teachers' baseline

characteristics, we examined the association between the coaching trajectory and baseline data. Importantly, the coaches did not have access to the belief and perception data, although they did have the dosage and fidelity data. When comparing the *low* and *moderate* classes, there were no significant differences on any of the study design variables (i.e., year in which the teacher participated in the study, the coach the teacher predominantly worked with, and the treatment condition), demographic data, or baseline implementation and beliefs and perceptions. It is possible that there was not enough power in the current study to detect the differences between the *low* and *moderate* classes; with additional teachers and greater power, the differences may have been significant. This implies that further research specifically examining low and moderate contact is warranted. Further, additional qualitative or quantitative data collection may be needed, including data from coaches about their decision making regarding the number of contacts a teacher receives as well as other teacher and classroom variables. For example, it is possible that teachers differed on other areas like classroom concerns such as whether student behavior was a challenge, their willingness to learn new skills, or the teacher's attitude about the value of the coaching support and the intervention.

Significant differences emerged across all domains when comparing the *low* and *high* classes. Specifically, teachers who had a greater need for support, as evidenced by their lower efficacy and higher burnout scores, were more likely to be in the *high* class, receiving intensive support over the course of the year. Although the coaches did not have access to these data elements, it seems likely that this reflects the skill of the coaches to assess the teachers' perceptions and emotions informally and to provide what was needed to the teachers based on their clinical and practical experience. Interestingly, coaches did have access to the teacher rubric scores, which were not related to class membership. Despite the teachers' ability to implement the game well, as assessed by the rubric quality rating, coaches may have been able to detect other areas of concern such as an unwillingness to implement the games regularly or high levels of stress. There were also some associations between study design elements and the receipt of a higher level of coach contacts. The added complexity of the integrated intervention (compared to PAX GBG only) may have prompted coaches and teachers to engage more with the coaching process, as was hypothesized. The differences in coaching support for teachers in cohorts/study year 1 and 3 may have related to the number of challenges faced by the schools, as compared to the year 2 schools.

In the final aim of the study, we examined the relationships between class membership and end-of-year data,

whereby differences in implementation dosage and beliefs and perceptions also emerged. Teachers in the *low* contact class were the most likely to demonstrate poor dosage when compared to the *moderate* and *high* contact classes, as measured by their lower implementation of games. This statistically significant difference was also present for the number of minutes played when comparing the teachers in the *high* to *low* class. It is important to note that the implementation dosage outcome was a summation of games and minutes across the entire school year. While there were no specific benchmarks established for the number of games that teachers would play across an entire year, coaches did collect weekly dosage data from teachers and teachers were given some general recommendations for daily use of the game. On the other hand, there were no end-of-year differences on the rubric assessing quality between coach contact classes, which was used explicitly to tailor the coaching on an ongoing basis, with the aim of minimizing implementation variability. As displayed in Table 1, rubric ratings on average were consistent across the year and were generally toward the higher end of the range (i.e., a score of three, where four was the highest possible score). It is possible that the lack of differences on the rubric arose from the greater restriction of range on this measure (i.e., 0–4) as compared to the dosage indicators, which reflected the cumulative number of games and minutes playing the PAX GBG. Teachers in the *low* contact class reported better perceptions at the end of the year as compared to the *moderate* contact class; teachers in the *moderate* class were more likely to have provided low ratings of organizational health and high ratings of depersonalization. Importantly, these differences were not present at baseline.

When considering that *low* contact teachers had higher baseline ratings of efficacy and less reported burnout than teachers in the *high* contact class, it seems plausible that *low* contact teachers did not perceive a need for the GBG and thus did not implement it at as frequently. By the end of the school year, the teachers in the *low* class were also more satisfied with the environment and felt less emotionally exhausted than those in the *moderate,* but not *high*, class. The teachers receiving a low level of support seemed to have a more generally positive experience and the fact that they had less frequent check-ins and coaching support may have led them to implement the program less regularly. Further, the nonsignificant differences between the *low* and *high* classes on beliefs and perceptions, which were present at baseline, also may reflect that high needs teachers who received intensive coaching supports benefitted. More research regarding this possible feedback loop and whether these *low* classrooms truly had fewer needs (e.g., as shown by student measures of behavior and achievement) is needed. In addition, research about the

impacts of varying levels of coaching on teachers' beliefs and perceptions is also needed. It is possible that teachers with a specific profile will most benefit from coaching.

Study Limitations

Although this study begins to highlight some relationships between beginning- and end-of-year data with coaching dosage, causal and directional conclusions cannot be drawn. More specifically, since coaching levels were not randomly assigned, we cannot make causal inferences regarding the impact of coaching trajectories on end-of-year teacher outcomes. Rather, it appears that the coaches employed a tailoring process, whereby they based their decisions about coaching supports on a range of actual data, and possibly some personal characteristics, in order to determine the level of supports (i.e., number of contacts) that they provided to the teachers. These data are from an RCT evaluating two interventions; it is possible that the coaching dosage and tailoring decisions made in the current study would not generalize to the coaching of other EBIs or instructional strategies.

There are also some measurement limitations to consider. Both dosage and beliefs and perceptions data were collected via teacher surveys, which may limit the conclusions; however, the implementation quality indicator did include assessment by an outside rater. Interestingly, the rubric quality indicator was not related to the class membership, whereas teacher reports were related. More research is needed on the outcome measures, including further validation regarding the 66th percentile cutpoint as an indicator of adequate implementation and teacher beliefs and perceptions. It is possible that the findings were sensitive to the cutpoint chosen (e.g., better distinction between the classes may have occurred at a higher threshold). The establishment of predetermined cutpoints for implementation dosage as well as quality that are associated with student outcomes would impact how coaches tailor their supports. Further data on classroom composition and student characteristics were not available but would likely expand our understanding of the coaching tailoring. In addition, a measure assessing the decision making of coaches as they proceeded through the process would have provided valuable insight and should be considered for inclusion in subsequent studies.

Finally, although the teacher and coach samples may seem small (i.e., three coaches and 210 teachers), the sample is relatively large for implementation studies within the coaching literature; in fact, many studies include far fewer teachers and just one or two coaches (for a review of the coaching literature, see Pas et al., 2014). Despite this fact, power concerns arising from the sample size may have limited our statistical testing. Further, we analyzed the data using GMM based on four waves of data that were not equally spaced.

Conclusions and Implications

In summary, the current study suggests that teachers who received a high degree of support generally reported more negative beliefs and perceptions at the start of the school year than those in the *low* contact class. At the end of the school year, teachers in the *low* contact class demonstrated the lowest GBG dosage, but also reported lower burnout and better school organizational health. Nevertheless, the coaching dosage was not significantly related to the observations of implementation quality, despite being associated with implementation dosage. These findings highlight the importance of examining variation in coaching supports, as they may be related to teacher, and possibly student, outcomes. The results of this study may also inform future RCTs testing coaching models as well as scale-up efforts of EBIs that include coaching. Taken together these findings suggest that coaches and implementation specialists should carefully attend to dosage indicators as well as teachers' beliefs and perceptions, either explicitly or implicitly, as they may be informative in tailoring of coaching supports.

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.

Baker, P. H. (2005). Managing student behavior: How ready are teachers to meet the challenge? *American Secondary Education, 33*, 51–64.

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.

Becker, K., Bradshaw, C. P., Domitrovich, C. E., & Ialongo, N. S. (2013a). Coaching teachers to improve the implementation quality of evidence-based programs: Linking coaching with fidelity in the PATHS to PAX Project. *Administration and Policy in Mental Health and Mental Health Services Research,*. doi:10.1007/s10488-013-0482-8.

Becker, K., Darney, D., Domitrovich, C., Keperling, J., & Ialongo, N. (2013b). Supporting universal prevention programs: A two-phased coaching model. *Clinical Child and Family Psychology Review, 16*, 213–228. doi:10.1007/s10567-013-0134-2.

Conduct Problems Prevention Research Group (CPPRG;. (2010). The effects of a multi-year randomized clinical trial of a universal social–emotional learning program: The role of student and school characteristics. *Journal of Consulting and Clinical Psychology, 78*, 156–168. doi:10.1037/a0018607.

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. doi:10.1002/pits.20452.

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. & Poduska, J. (2008). *The social–emotional learning efficacy scale*. Unpublished technical report.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82,* 405–432. doi:10.1111/j.1467-8624.2010.01564.x.

Dusenbury, L., Branningan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18,* 237–256.

Embry, D., Staatemeier, G., Richardson, C., Lauger, K., & Mitich, J. (2003). *The PAX good behavior game* (1st ed.). Center City, MN: Hazelden.

Gottfredson, G. D., Jones, E. M., & Gore, T. W. (2002). Implementation and evaluation of a cognitive-behavioral intervention to prevent problem behavior in a disorganized school. *Prevention Science, 3,* 43–56.

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., Cook, E., & Quamma, J. (1995). Promoting emotional competence in school-aged children: The effects of the PATHS curriculum. *Development and Psychopathology, 7,* 117–136.

Hicks, T. B., Shahidullah, J. D., Carlson, J. S., & Palejwala, M. H. (2014). Nationally Certified School Psychologists' use and reported barriers to using evidence-based interventions in schools: The influence of graduate program training and education. *School Psychology Quarterly, 29,* 469–487. doi:10.1037/spq0000059.

Hoagwood, K. E., & Burns, B. J. (2005). Evidence-based practice, part II: Effecting change. *Child Adolescent Psychiatric Clinics of North America, 14*(2), xv–xvii.

Hoagwood, K. E., Olin, S. S., Kerker, B. D., Kratochwill, T. R., Crowe, M., & Saka, N. (2007). Empirically based school interventions targeted at academic and mental health functioning. *Journal of Emotional & Behavioral Disorders, 15,* 66–92. doi:10.1177/10634266070150020301.

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.

Ingersoll, R. M., & Smith, T. M. (2003). The wrong solution to the teacher shortage. *Educational Leadership, 60*(8), 30–33.

Kaiser, A., & Cross, F. (2011). *Beginning teacher attrition and mobility: Results from the first through third waves of the 2007–08 Beginning Teacher Longitudinal Study*. Washington, DC: U.S. Department of Education. Retrieved from http://nces.ed.gov/pubs2011/2011318.pdf

Kellam, S. G., Poduska, J. M., Ialongo, N. S., Wang, W., Toyinbo, P., Petras, H., et al. (2008). 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. doi:10.1016/j.drugalcdep.2008.01.004.

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

Lo, Y., Mendell, N., & Rubin, D. (2001). Testing the number of components in a normal mixture. *Biometrika, 88,* 767–778.

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. doi:10.14221/ajte2008v33n4.3.

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Occupational Behavior, 2,* 99–113.

Maslach, C., Jackson, S. E., & Leiter, M. P. (1997). Maslach burnout inventory: Third edition. In R. J. Wood (Ed.), *Evaluating stress: A book of resources* (pp. 191–218). Scarecrow Education: Lanham, MD.

Muthén, B. O. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *The SAGE handbook of quantitative methodology for the social sciences* (pp. 345–368). Thousand Oaks, CA: Sage.

Muthén, L. K., & Muthén, B. O. (1998–2013). *Mplus user's guide*. Los Angeles, CA: Muthén and Muthén.

Nagin, D. S. (2005). *Group-based modeling of development over the life course*. Cambridge, MA: Harvard University Press.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14,* 535–569. doi:10.1080/10705510701575396.

O'Connell, M. E., Boat, T., & Warner, K. E. (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities*. Washington, DC: The National Academies Press. Committee on the Prevention of Mental Disorders and Substance Abuse Among Children, Youth, and Young Adults: Research Advances and Promising Interventions.

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* (2nd ed., pp. 255–268). New York: Springer.

Pas, E. T., & Newman, D. L. (2013). Teacher mentoring, coaching, and consultation. In J. A. C. Hattie & E. M. Anderman (Eds.), *International handbook of student achievement* (pp. 152–154). New York: Routledge Publishing Company.

Ramaswamy, V., Desarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science, 12,* 103–124.

Reinke, W. M., Stormont, M., Herman, K. C., Puri, R., & Goel, N. (2011). Supporting children's mental health in schools: Teacher perceptions of needs, roles, and barriers. *School Psychology Quarterly, 26,* 1–13. doi:10.1037/a0022714.

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 and Behavior, 30,* 375–391.

Ringwalt, C. L., Vincus, A. A., Hanley, S., Ennett, S. T., Bowling, J. M., & Haws, S. (2011). The prevalence of evidence-based drug use prevention curricula in US middle schools in 2008. *Prevention Science, 12,* 63–69. doi:10.1007/s11121-008-0112-y.

Schaffer, K., Rouiller, S., Embry, D., & Ialongo, N. (2006). *The PAX good behavior game implementation rubric*. Unpublished technical report. Johns Hopkins University.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461–464.

Siebert, C. J. (2005). Promoting preservice teachers' success in classroom management by leveraging a local union's resources: A professional development school initiative. *Education, 125*, 385–392.

Spoth, R., Rohrbach, L. A., Greenberg, M., Leaf, P., Brown, C. H., Fagan, A., et al. (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*, 319–351. doi:10.1007/s11121-012-0362-6.

U.S. Department of Education. (2012). *Prevalence and implementation fidelity of research-based prevention programs in public schools: Final report*. Washington, DC: U.S. Department of Education. Retrieved from http://www2.ed.gov/rschstat/eval/other/research-based-prevention.pdf

Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., et al. (2008). Bridging the gap between prevention research and practice: The interactive systems framework for dissemination and implementation. *American Journal of Community Psychology, 41*(3–4), 171–181. doi:10.1007/s10464-008-9174-z.

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*, 22–33. doi:10.1007/s12310-011-9067-4.

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.

Wilson, S. J., & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior: Update of a meta-analysis. *American Journal of Preventive Medicine, 33*, 130–143.