ORIGINAL PAPER

The Impact of Working Alliance, Social Validity, and Teacher Burnout on Implementation Fidelity of the Good Behavior Game

Joseph H. Wehby · Daniel M. Maggin ·
Tara C. Moore Partin · Rachel Robertson

Abstract This study examined the association of three teacher-specific factors on their level of treatment implementation of an evidence-based classroom management strategy. The three factors were (a) the teacher–coach alliance, (b) teacher ratings of the intervention's social validity, and (c) teacher burnout. Results indicated that both teacher–coach alliance and social validity were uniquely associated with treatment implementation. In addition, a model with all predictors revealed that the teacher–coach alliance had the largest relation with treatment implementation. It was also found that the reported level of teacher burnout moderated the relation between teacher–coach alliance and procedural fidelity. Specifically, teacher burnout was found to have a negative relation to treatment implementation at low levels of teacher–coach alliance but no relation to treatment implementation at high levels of teacher–coach alliance. These findings suggest that the teacher–coach relationship may not only play a direct role in enhanced treatment implementation for school-based programs but also may mitigate the effects of teacher burnout.

Introduction

The implementation of evidence-based practices in schools has received considerable attention in recent years (e.g., Carnine, 1995; Singh & Oswald, 2004). Efforts by the Institute of Education Sciences (IES) and other agencies to identify, validate, and bring effective treatments to scale have led to additional concerns regarding the capacity of school personnel to implement programs with integrity (Walker, 2004). Until recently, treatment integrity has primarily concerned intervention researchers interested in ensuring that observed treatment effects were, in fact, the result of the program being delivered. That is, treatment integrity data are used within research to facilitate interpretation of outcomes by ensuring the intervention was executed as designed (Shadish, Cook, & Campbell, 2002). For instance, consider the case in which treatment integrity is low and effects are not present, a reasonable conclusion might be that improved implementation would result in higher effects. Conversely, if adequate treatment integrity is observed and effects were still not present, the interpretation would be that the intervention was ineffective. The importance of treatment integrity is made even more apparent by research indicating that the magnitude of treatment effect is often associated with the level of implementation (Perepletchikova & Kazdin, 2005). Unfortunately, there is also evidence indicating that

J. H. Wehby
Vanderbilt University, Nashville, TN, USA

D. M. Maggin (✉)
School of Education, University of Illinois at Chicago,
3343 EPASW, 1040 W. Harrison M/C 147, Chicago,
IL 60607, USA
e-mail: dmaggin@uic.edu

T. C. Moore Partin
University of Tennessee-Knoxville, Knoxville, TN, USA

R. Robertson
Duqesne University, Pittsburgh, PA, USA

treatment integrity, and subsequent effects, often deteriorate when moving from more to less controlled environments (Hulleman & Cordray, 2009). As such, the positive effects observed during the validation process of an intervention are less likely to be replicated in "real-world" settings and the issue of treatment integrity necessarily evolves from interpretation to transportability (Greenberg, Domitrovich, & Bumbarger, 2001).

The failure of many educational interventions to translate from research to practice has led to several investigations regarding the factors impeding program adoption and implementation (Fixen, Blase, Naoom, & Wallace, 2009). A majority of this literature has focused on organizational and policy features within schools to support the use and implementation of effective practices (e.g., Deshler, 2003; Elmore, 1996). In contrast, few studies have examined the influence of teacher-specific factors on levels of treatment implementation despite the fact that many school-based treatments rely on classroom teachers for delivery (Greenberg et al., 2001; Han & Weiss, 2005). As such, the present study is an attempt to extend the findings of previous research by identifying teacher-level factors related to implementation of an evidence-based classroom management program. The factors used within the analyses represent a subset of teacher-specific variables identified by Han and Weiss thought to impact the implementation and adoption of evidence-based practices. Specifically, the relationship between teacher implementation of the Good Behavior Game (GBG; Barrish, Saunders, & Wolf, 1969) and (a) teacher perceptions of the quality of the coach–teacher relationship, (b) the perceived effectiveness and utility of the intervention, and (c) a measure of teacher stress were investigated. In the following sections, we provide descriptions of the factors considered within the present study. These brief reviews will be used to demonstrate the utility of the research being conducted as well as identifying areas in which the present study extends the current literature.

Working Alliance

The acknowledgement that empirically supported interventions are often not used with integrity in applied contexts has led to the development of several methods designed to increase the adherence of school personnel to program components (Fixen et al., 2009). Among the most widely adopted methods for ensuring that programs are implemented with integrity is the school-based coach (Joyce & Showers, 1995). School coaches are typically charged with working directly with teachers to assist with the implementation of specific programs by providing performance feedback, answering questions regarding program components, and assisting with developing individual or class delivery plans (Yopp et al., 2011). While school-coaching has been shown to positively influence levels of treatment implementation by teachers and other school personnel (Kretlow & Bartholomew, 2010), the effects of coaching might ultimately be enhanced or inhibited by additional factors. One example of a factor that might moderate the effects of coaching on treatment integrity is the quality of the relationship developed between the teacher and coach charged with delivering the feedback. That is, teachers might be more likely to implement a strategy based on the perceived quality of the relationship with their coach. This construct is closely related to therapeutic alliance, which is a concept drawn from the psychotherapy literature and refers to the collaborative and affective bond between therapist and patient (Martin, Garske, & Davis, 2000). Several research syntheses examining the impact of therapeutic alliance on treatment outcomes have demonstrated a moderate but consistent association between alliance and implementation of program components (e.g., Horvath & Luborksy, 1993; Horvath & Symonds, 1991; Martin et al., 2000). To our knowledge, there are only a few examples of working alliance measures applied to the coach–interventionist relationship in school-based research (e.g., Bierman, 2002; Seeley et al., 2009). Reports of coach–teacher alliance in these reports have been used solely for descriptive purposes to illustrate the quality of implementation. None of these examples have considered the impact of alliance on the level of treatment fidelity exhibited by teachers or other school-based interventionists. In terms of implementation research, however, the coach–teacher relationship might prove to be a critical construct in need of further investigation.

Social Validity

Social validity refers to an estimate of the importance, effectiveness, appropriateness, and satisfaction an implementer has with a particular intervention (Kennedy, 2005; Wolf, 1978). Many authors have surmised that issues related to social validity are among the most important predictor of implementation (Elliott, 1988; Lentz, Allen, & Erhardt, 1996; Reimers, Wacker, & Koeppl, 1987). Witt and Elliott (1985), for instance, hypothesized that treatment acceptability would be positively related to both the initiation and integrity of practices. Despite the prevalent conceptual support, few attempts have been made to empirically demonstrate the impact of social validity on treatment integrity (Sterling-Turner, Watson, & Moore, 2002). In addition, most of the empirical research investigating the association between social validity and treatment integrity has been conducted using analog approaches in which treatment evaluators are not the ones charged with implementing the practice (Gresham & Lopez, 1996; Sterling-Turner et al., 2002). Interestingly, the results of

studies investigating the direct relation between social validity and treatment integrity have been mixed. Initial findings seemed to indicate that social validity was, in fact, associated with greater implementation (Allinder & Oats, 1997). However, each of these studies relied on implementation measures using teacher report. As a result, it is unclear the extent to which these relations might be confounded by method variance. To improve on these and other methodological concerns, Sterling-Turner et al. (2002) conducted an experimental evaluation of the impact of social validity on measures of treatment implementation and determined that there was no relation between social validity and treatment integrity. However, this study took place under highly controlled circumstances, and the authors were uncertain as to whether their findings would generalize to clinical settings. Therefore, additional research is needed to better understand the association between treatment integrity and social validity within the context of applied coaching interactions.

Educator Burnout

Both Han and Weiss (2005) and Greenberg et al., (2001) have advocated for investigators to consider the relation between educator stress and school-based treatments. Although few studies have examined the relationship between educator burnout and treatment implementation directly, there is some evidence to suggest that higher levels of stress might impact the ability and willingness of teachers to use evidence-based procedures (Ransford, Greenberg, Domitrovich, Small, & Joacobson, 2009). In addition to evidence that suggests burnout is related to greater rates of teacher turnover, school absences, and negative interactions with students (Cooley & Yovanoff, 1996; Guin, 2004; Schwab, Jackson, & Schuler, 1986), teachers reported more negative attitudes toward the implementation of a novel school practice as a function of burnout (Evers, Brouwers, & Tomic, 2002). Han and Weiss noted that, to date, there has been no evidence supporting a direct link between educator burnout and treatment implementation. However, the relation between educator burnout and other negative teacher outcomes (i.e., attrition; absenteeism; negative interactions) suggests that burnout might influence teacher willingness to incorporate new approaches regardless of the evidence supporting the programs' effectiveness.

Purpose and Contribution of Study

The purpose of the present study was to examine the relations between (a) working alliance, (b) social validity, and (c) educator burnout with implementation fidelity of the GBG (Barrish et al., 1969). The present study extends previous work in four ways: First, to the best of our knowledge, this is the first attempt to examine the impact of working alliance on teachers' implementation fidelity. The association between the perceived relationship quality of coaches and teachers might have important ramifications for both research and practice. Second, the relation between social validity and treatment integrity will be analyzed within an applied context. As discussed before, many have theorized that treatment acceptability will influence the level and quality of implementation (e.g., Witt & Elliott, 1985), but few studies have directly examined this relation with independent observations of implementation. Third, some have recently acknowledged that issues of occupational stress might impact teacher implementation (Greenberg et al., 2001; Han & Weiss, 2005). Research supporting this notion is largely anecdotal, and the present study directly investigates the association between teacher burnout and levels of implementation. Fourth, in addition to considering these variables individually, an attempt will be made to delineate which of these variables contributes most to the measure of treatment integrity.

Method

The current study was conducted within the context of a larger, federally funded project designed to test a treatment package to increase the academic achievement and prosocial behaviors of children with and at risk for emotional and behavioral disorders (EBD). Teacher participants were randomly assigned to either a treatment or control group. Those teachers receiving the intervention were trained to incorporate the GBG and audio self-monitoring into their daily classroom instruction. Specifically, teachers were asked to use the GBG for a minimum of 20 min each day during language arts instruction. In addition, teachers took 5-min samples of instruction and tracked their rates of praise, reprimands, and academic prompts delivered to students. Implementation assessments were conducted through direct observation by the assigned coach. Following the completion of an observation session, the coaches provided graphic feedback on teacher fidelity to intervention components.

Teacher Recruitment Procedures

Following institutional approval of the project, school district personnel were asked for permission to contact principals for participation in the study. A meeting was held with each interested principal to provide further detail regarding the purpose, procedures, and outcomes of participating in the study. For those principals who remained interested in participating in the study, the general and special education teachers were informed of the study's purpose and were

asked to consider taking part in the project. All teachers were notified that agreeing to participate would not necessarily result in immediate training on intervention components and that the timing of training was subject to random assignment of their school to treatment or control. However, teachers were ensured training to intervention components at some point over the course of the project. It should be noted that all special education teachers contacted to participate consented while only those general education teachers that expressed interested were contacted to obtain consent. These procedures were followed in both the first and second year of the project.

Teacher Participant Description

Across the study's 2 years, a total of 163 teachers were randomly assigned to receive training in a multi-component classroom management program with ongoing coaching support or to an assessment-only control group. The present study focused on the subset of 82 teachers who were randomized to the treatment group. Of these 82 teachers, nine were missing either treatment integrity or alliance data, and the final sample therefore included 73 elementary school teachers. A total of 56 intervention teachers were recruited for the first year while 17 intervention teachers began during the second year of the study. The sample included 48 general education teachers and 25 special education teachers. The general education classrooms were structured for typically developing students whereas the special education classrooms were structured for children with EBD. Structural differences between these settings included (a) lower teacher-student ratio in self-contained environments (typically 8:1)

in the special education classrooms and (b) the presence of a classroom paraeducator in the special education classrooms to assist the teacher.

Table 1 provides descriptive statistics for teacher demographic data. Group comparisons were conducted between general and special education teachers on these demographic variables. Results of these Chi-square analyses revealed one significant difference between the groups with, not surprisingly, special educators more likely to have a special education credential than general education teaches. Because general and special education teachers were both sampled, preliminary analyses were conducted to compare teachers from the two different settings on each of the dependent and independent variables to determine whether general and special education teachers could be pooled into a single sample. A series of one-way ANOVAs were used to compare group means on each of the variables. Results indicated that the setting (i.e., general or special education) was not significantly related to (a) the percent of steps implemented (F [1, 72] = 2.32, p = 0.13), (b) social validity ratings (F [1, 72] = 0.00, p = 0.99), (c) alliance ratings F [1, 72] = 0.07, p = 0.79), or (d) teacher burnout F [1, 72] = 1.38, p = 0.24). Therefore, the teachers were pooled into a single group for later analyses.

Coach Description

A total of 12 coaches were used over the course of the study with the 73 teachers. The coaches' role was to provide teachers with (a) training on intervention components, (b) project resources (e.g., reinforcers, GBG Board), (c) feedback regarding the quality of the teacher's

Table 1 Teacher demographics by setting and total sample

Variable	Level	General (n = 48)	Special (n = 25)	Total (n = 73)
Gender n (%)	Male	5 (10.42%)	4 (16.00%)	9 (12.33%)
	Female	43 (89.58%)	21 (84.00%)	64 (87.67%)
Race n (%)	AA	12 (25.00%)	10 (40.00%)	22 (30.14%)
	EA	36 (75.00%)	14 (56.00%)	50 (68.49%)
	Other	0 (0.00%)	1 (4.00%)	1 (1.37%)
Grade n (%)	Kindergarten	10 (28.83%)		10 (28.83%)
	First	11 (22.92%)		11 (22.92%)
	Second	12 (25.00%)		12 (16.44%)
	Third	11 (22.92%)		11 (15.07%)
	Fourth	4 (8.22%)		4 (5.48%)
	Multiple		25 (100.00%)	25 (34.25%)
Education n (%)	Bachelor's	27 (56.25%)	15 (60.00%)	42 (57.54%)
	Master's	21 (43.75%)	10 (40.00%)	31 (42.47%)
Certification n (%)	General	34 (70.83%)	3 (12.00%)	37 (50.68%)
	Special	14 (29.17%)	22 (84.00%)	36 (47.95%)
Years experience M (SD)		12.30 (11.88)	12.58 (10.64)	12.39 (11.41)

implementation, (d) assistance with specific issues related to classroom management, (e) troubleshooting issues in regards to intervention implementation, and (f) a liaison between teachers and other project staff. In addition, coaches were responsible for collecting questionnaires from the teachers, including measures of treatment implementation. Each coach received initial training on implementing treatment components, conducting feedback sessions, and completing project forms including measures of fidelity.

Feedback regarding teachers' implementation of the GBG was typically provided to teachers by the coaches on a biweekly basis. These meetings usually took place during the teacher's planning period or other non-instructional time that was convenient for the teacher (i.e., after school). Issues related to any part of the intervention or other classroom management topics were discussed during these meetings as well. Demographic and dosage data for coaches are provided in Table 2. Dosage refers to the total number of meetings held between the teacher and coach regarding project issues.

Good Behavior Game

The GBG has substantial evidence supporting its efficacy for improving the classroom behavior of students at risk for a variety of challenging behaviors disruptive (e.g., Davies & Witte, 2000; Johnson, Turner, & Konarski, 1978; Medland & Stachnik, 1972). Further, the efficacy of the GBG program has been replicated across different grade levels, types of students, and settings, and has consequently led to a number of federal agencies to deem this program as a 'best' practice

Table 2 Coach demographic characteristics

Variable	Mean (SD)	Range	n (%)
Age	31.44 (10.92)	22–58	
Female			10 (62.50)
Bachelor's			6 (37.50)
Master's			9 (56.25)
PhD			1 (6.25)
Caucasian			13 (81.25)
Teaching experience[a]			8 (50.00)
Year 1 integrity observations	4.55 (2.18)	1–12	
Year 2 integrity observations	9.85 (5.20)	2–25	
Year 1 coach dosage[b]	5.91 (1.99)	3–12	
Year 2 coach dosage[b]	10.18 (5.03)	2–25	

[a] Measured as dichotomous variable with "yes" indicating teaching experience

[b] Measured as the total number of times the teacher met with the coach

(Embry, 2002; Tingstrom, Sterling-Turner, & Wilczynski, 2006). In the most basic definition, the GBG is a prescriptive interdependent group-contingency program based on the principles of behavior modification. Within the current version of the program, teachers developed a set of operationally defined inappropriate classroom behaviors and identified a set of appropriate reinforcers to be used as rewards with the guidance of a research coach. The class was then split into teams and subsequently competed as a group for prizes, privileges, and activities. Tally marks were placed on a game board for a team when a disruptive behavior of any team member occurred. If the tally marks for a team remained below a pre-determined threshold number (e.g., eight) by the end of the game, the team won. All teams were eligible to win if their tallies did not exceed the preset threshold number. Full implementation of the GBG required executing a total of 18 steps. Basic steps of the GBG included (a) announcing the game, (b) reminding students of the configuration of each team, (c) reviewing the classroom rules, and (d) announcing the threshold number of behaviors required to win the game. Adjustments to the threshold or length of game time were made if teams won consistently or were not winning consistently enough to successfully modify behavior. For the present study, teachers were recommended to begin with a threshold of eight disruptive behaviors and a game time of about 30 min. However, it should be noted that the flexibility of the GBG is among its greatest advantages as an evidence-based practice. In other words, teachers were able to judge appropriate thresholds and game lengths based on their classroom needs.

Measures

Treatment Implementation

An observer checklist developed for this study was used to measure teacher adherence to the GBG procedures. The 18 items on the checklist were based directly on the procedures described in the GBG manual (Dolan, Turkkan, Werthamer-Larsson, & Kellam, 1989). Raters placed a check mark into the corresponding "yes" or "no" box to indicate if that particular step was observed or not. Example items include (a) announce game before beginning, (b) explain rule violation process, and (c) record each team's performance on data sheet. These data were collected multiple times throughout the course of the project and subsequently provided to teachers as feedback on their progress and level of implementation. On average, procedural fidelity was collected 4.61 times for teachers participating in the first year of the study and 8.76 times for teachers participating in the second year. The ratio between average number of times integrity was collected and weeks of intervention was approximately 2:1 for both years. The

estimates used for the present analyses were based on data from a single time point taken toward the end of the intervention period because it was at this point that the sample of teachers demonstrated the greatest implementation integrity. Moreover, teachers were blinded to this fidelity check whereas in all other cases, they were not. As such, these data were considered more objective than other administrations of the fidelity checklist. For teachers participating in both years of the study, an average of blind fidelity ratings was calculated across Year 1 and Year 2 checks. The reliability of the fidelity instrument was assessed by averaging the test–retest correlations across all adjacent administrations of the fidelity protocol and resulted in a 0.78 average coefficient. In addition, an alpha coefficient was computed for the specific fidelity checklist used for data analysis which revealed good internal consistency and inter-item covariance. Specifically, the internal consistency was 0.86 with an inter-item covariance of 0.39.

Alliance Ratings

The teacher–coach alliance scale was developed by members of the National Behavior Rating Coordination Center (NBRCC), which was established to assist four behavior research centers at universities across the United States. The scale was designed to measure the perceived quality of the teacher–coach relationship from the perspective of the teacher. Each of the ten items on the scale was rated on a 5-point Likert scale (never = 1, seldom = 2, sometimes = 3, often = 4, always = 5). Example items include (a) the teacher/coach and I agree on what the most important goals for intervention are; (b) I feel confident of the teacher/coach's ability to help the situation; and (c) the teacher/coach is approachable. Administration time was approximately 3 min. For the present study, the scale was administered to teachers and coaches toward the end of the intervention for both Year 1 and Year 2. Teachers were given the scale as part of a packet of project measures for the teacher to complete on their own time. In many cases, but not all, the coach was the individual delivering and receiving the materials. However, teachers completed the materials at a time when the coach was not present. Items were scaled using a percentage of total possible points. Specifically, the sum of all the items for which responses were provided was computed. This sum was then divided by the total possible points that the respondent could have received. For instance, if a respondent did not complete a particular item this item was not used to compute the percentage of total points. Since the study was longitudinal and a majority of teachers and coaches participated in both Year 1 and Year 2, scores across both years were averaged for analyses. Preliminary analyses of teacher responses on the alliance scale indicated strong psychometric properties.

The alpha coefficient revealed an internal consistency rating of 0.96 with an average inter-item covariance of 0.57 for data aggregated across the whole sample. In terms of the present sample, therefore, the scale proved to have strong reliability.

Social Validity Ratings

A researcher developed rating form was used to measure social validity. The scale had a total of 13 items that surveyed teacher perception of the (a) effectiveness, (b) fit, and (c) burden of the GBG for the teacher and their classroom. Each of the items was rated on a 5-point Likert scale (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree). Items were developed to estimate teacher's perspectives on the importance, effectiveness, appropriateness, and satisfaction with the GBG. Example items included (a) I plan to use the GBG in my classroom in the future, (b) the GBG was a good fit for my classroom, and (c) the addition of the GBG has improved behavior in my classroom. The social validity scale was administered at the end of the intervention period with a packet of project materials. Scores were averaged across both Year 1 and Year 2 for teachers that were in the project both years. Internal reliability estimates of the social validity scale revealed consistency between items with an alpha coefficient of 0.94 and inter-item covariance of 0.45.

Teacher Burnout

The Emotional Exhaustion subscale of the Maslach Burnout Inventory Educators Survey (MBI-ES; Maslach, Jackson, & Schwab, 1986) was used to measure levels of teacher burnout. The MBI-ES consists of 22 total items. Each of the 22 items was rated on a 6-point Likert scale (0 = never; 1 = a few times a year or less; 2 = once a month; 3 = a few times a month; 4 = once a week; 5 = a few times a week; 6 = everyday). Although the MBI-ES has three subscales of burnout including Emotional Exhaustion, Depersonalization, and reduced Personal Accomplishment, correlations among the Depersonalization and Personal Accomplishment subscales revealed small and non-significant associations with measures of integrity. The range of absolute values of these correlations was 0.01–0.16 with none being statistically significant. In contrast, the Emotional Exhaustion subscale was moderately and significantly correlated with the treatment integrity variables included in analyses. Therefore, the Emotional Exhaustion subscale was included in the present analyses. The Emotional Exhaustion subscale has a total of nine items designed to measure the degree of emotional and physical fatigue experienced by the educator. Example

items from this subscale included (a) I feel emotionally drained from work; (b) I feel like I'm at the end of my rope; and (c) I feel burned out from my work. For the present study, the MBI-ES was administered prior to the intervention for both years of the project. These pretest scores were used in data analyses. If teachers participated in both years of the study, their scores prior to the first year were used. Scores on the Emotional Exhaustion subscale were derived using total sums on the relevant items. For the present sample, the Emotional Exhaustion scale of the MBI demonstrated strong psychometric properties with an alpha coefficient of 0.91 and inter-item covariance of 0.53.

Preliminary Data Analysis

The dependent variable for all models analyzed included the percent of steps implemented during the official fidelity check as the dependent variable. In addition, preliminary analyses were conducted to determine whether there were differences between coaches regarding teacher implementation. These analyses were conducted in order to determine whether coaches should be included as a random effect in subsequent models. An unconditional multilevel model was tested, which, in effect, was a one-way ANOVA with random intercepts (Raudenbush & Bryk, 2002). This model was used to (a) calculate the level of dependence between teachers with the same coach through computation of the intraclass correlation (ICC) and (b) test the significance of the ICC. The ICC represents the correlation on the outcome variable (i.e., treatment integrity) between two randomly drawn individuals within the same cluster (Snijders & Boskers, 1999). Results of these analyses demonstrated that teacher fidelity ratings were, in fact, dependent on coach assignment. The ICC for steps implemented was 0.48 ($\chi^2 = 79.97$, $p < 0.001$).

The large and statistically significant ICCs indicated dependence among treatment implementation ratings of teachers with the same coach. In order to account for the dependence among observations, a multilevel random coefficient regression model (Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002) was used to estimate the variances associated with each predictor variable on treatment implementation with coach assignment included as a random effect. Due to the low number of coaches, multilevel modeling is a rather conservative approach for the present analyses. However, multilevel modeling was chosen over alternative methods (e.g., robust standard error adjustment) because it (a) corrects for the ICC and (b) matches the degrees of freedom to the number of clusters. This reduces the amount of variability in the predictors and residuals and leads to more conservative point estimates. Additional analyses were conducted using the Huber-White adjustment to standard errors. Results

associated with this approach differed slightly with the multilevel model. Specifically, the burnout predictor was significant with the standard error adjustments and not with the multilevel model. More detailed results of these analyses can be obtained by contacting the first author.

Results

Descriptive and Correlational Analyses

Means, standard deviations, ranges, and correlations for the variables used in this study are presented in Table 3. Correlational analyses were conducted to determine the strength of association between the variables to be included in subsequent analyses. In terms of correlations among predictor variables and the measure of treatment implementation, the social validity scale displayed moderate correlations with both. The alliance and burnout scales had small to moderate associations with the measures of treatment integrity. Notably, the burnout scale correlations had a negative association with both integrity measures. A moderately high correlation was observed between the alliance scale and the social validity measure. Both of these measures had negative, though non-significant association with the burnout scale. All other correlations were significant at the 0.05 level or lower.

Bivariate Relationships

A series of hierarchical regression models with coach assignment used as a random effect were tested to determine the relation between each predictor variable (i.e., alliance, social validity, and burnout) and levels of treatment implementation (i.e., percentage of steps implemented). These results are presented in Table 4. All predictor variables were centered on their respective means to facilitate interpretation. In addition to estimation of regression coefficients, the amount of variance explained

Table 3 Means, standard deviations, and correlation matrix of treatment fidelity and predictor variables

Variable	1	2	3	4
1. Steps implemented	–			
2. Working alliance	0.26	–		
3. Social validity	0.38	0.59	–	
4. Educator burnout	−0.27	−.13[a]	−.20[a]	–
Mean	65.60	90.53	65.99	18.89
Standard deviation	32.38	11.92	12.02	12.15
Range	0–100	22–87	37–100	0–45

[a] Represents correlations that were not significant at the 0.05 level. All other correlations were significant with p values <0.03

Table 4 Summary of bivariate regression analyses for variables predicting percentage of steps implemented for the good behavior game

Variable	β	SE β	R^2	p
Alliance	0.85	0.13	0.17	<0.01*
Social validity	0.76	0.23	0.08	<0.01*
Burnout	−0.26	0.28	0.04	0.35

* Indicates statistical significance at the $p < 0.05$ level

by each variable was computed using the formula provided by Snijders and Bosker (1999, p. 103). The percent of explained variance is reported as an R^2 statistic. Analyses concerning the measure of percentage of steps implemented revealed that teacher–coach alliance and social validity had a uniquely significant relationship with the dependent variable. Specifically, the alliance scale ($R^2 = 0.17$, $p < 0.01$) individually accounted for about 17% of the variance in the number of steps implemented. Although the predictive utility of social validity ($R^2 = 0.08$, $p < 0.01$) was not as strong, it was still found to have a significant association with percentage of steps implemented. The burnout scale was found to have a negative association with the percentage of steps implemented. However, this relationship was not empirically validated ($p = 0.35$). Since all variables were centered prior to data analysis, the interpretation of individual coefficients is based on standard scores. In other words, a standard deviation increase in teacher-coach alliance was associated with more than three-quarters of a standard deviation increase in treatment integrity. Similarly, a standard deviation increase in social validity rating was associated with about a three-quarter standard deviation increase in percentage of steps implemented.

Full Model Analyses

Following analyses considering individual predictors, hierarchical multiple regression analysis was used to estimate (a) the combined effect of the predictors and (b) the unique effect of each predictor on treatment implementation

when all other variables are held constant. Results of this set of analyses are presented in Table 5. In terms of the percentage of steps implemented, the full set of predictors explained about 24% of the variance with a comparison of the deviance statistics revealing a significant improvement in model fit with the additional parameters ($\chi^2 = 48.68$, $p < 0.01$). Coach alliance was found to be the only variable that had a statistically significant relationship with percentage of steps implemented. In contrast, both social validity and burnout had no unique effect. According to this analysis, a standard deviation increase in coach alliance was associated with more than half a standard deviation increase in the number of steps implemented after controlling for social validity and teacher burnout.

Full Model with Interaction Terms

A final set of analyses was conducted with the full model and a series of interaction terms tested. Two-way interactions between alliance and burnout, alliance and social validity, and social validity and burnout were tested. The results of these analyses are presented in Table 6 and revealed a significant interaction between alliance and teacher burnout. Specifically, the level of educator burnout was found to moderate the association between coach alliance and treatment fidelity. The test of the full model plus interaction term for percentage of steps implemented revealed that the interaction was significant ($p = 0.02$). Simple regression equations and tests were conducted following this significant finding. The graph of this relationship is presented in Fig. 1. Teachers were classified into those that reported high- (i.e., one standard deviation above), and low (i.e., one standard deviation below) levels of coach alliance with the impact of burnout plotted against percentage of steps implemented for each of these groups.

Discussion

Increased attention has been given to issues surrounding the implementation and sustainability of evidence-based

Table 5 Summary of full hierarchical regression analysis for variables predicting percentage of steps implemented of the good behavior game

Variable	β	SE β	p
Intercept	67.91	6.14	<0.01*
Alliance	0.59	0.19	<0.01*
Social validity	0.41	0.28	0.15
Burnout	−0.25	0.26	0.36

* Indicates statistical significance at the $p < 0.05$ level

Table 6 Summary of full hierarchical regression model with alliance by burnout interaction term predicting percentage of steps implemented for the good behavior game

Variable	β	SE β	β	p
Intercept	67.30	3.58		<0.01*
Social validity	0.19	0.36	0.07	0.61
Alliance	0.56	0.39	0.20	0.16
Burnout	−0.54	0.29	−0.20	0.07
Alliance × burnout	0.06	0.03	0.26	0.03*

* Indicates statistical significance at the $p < 0.05$ level

Fig. 1 Interaction of educator
burnout moderating the relation
between teacher implementation
of the good behavior game and
working alliance with coach
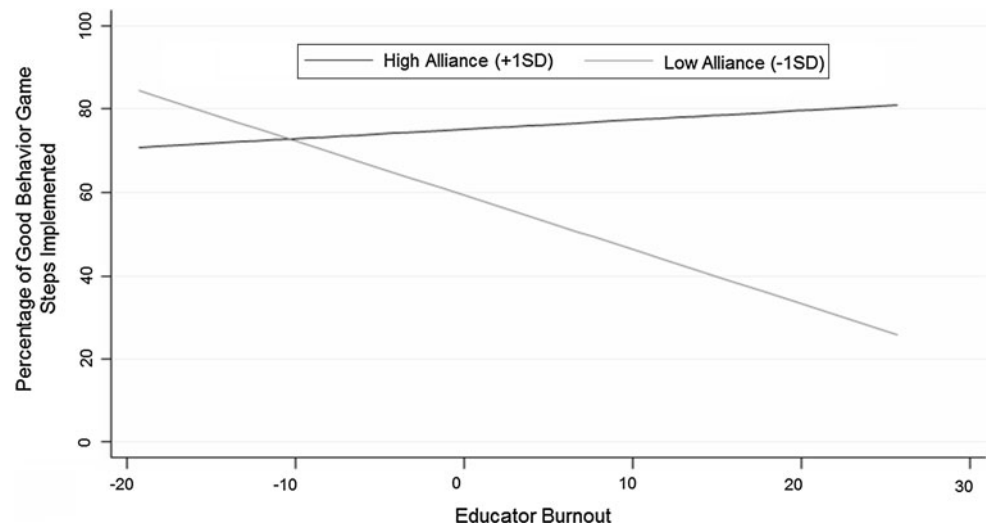
practices in schools in recent years (e.g., Elmore, 1996;
Fitzpatrick & Knowlton, 2009). This study explored the
unique and combined effects of three factors theorized to
influence teacher's procedural fidelity to an evidenced-
based classroom management program (Han & Weiss,
2005). Results of the bivariate analyses revealed that tea-
cher ratings of social validity accounted for the greatest
amount of variance in the number of steps implemented
with working alliance also having a moderate association.
In addition, the extent of educator burnout was negatively
related to the procedural fidelity of teachers. Notably,
educator burnout had a small relation with working alliance
and social validity though these correlations were found
not to be significant. The second step in the data analysis
was to determine the relation between each independent
variable on treatment implementation while controlling for
the other predictors. These analyses indicated that working
alliance was the only variable to have a unique effect on
implementation controlling for social validity and educator
burnout. The third and final step of data analysis was to
determine whether an interaction effect was present within
the data. This set of analyses revealed that the association
between working alliance and treatment implementation
was moderated by educator burnout for teachers with low
levels of alliance. This suggests that good coach–teacher
relationships may mitigate the potentially negative effects
of educator burnout on treatment adherence.

Extensions to the Literature

The results of this study provide a basis for further research
into the factors that moderate levels of procedural fidelity
for teachers charged with executing evidence-based class-
room interventions. Recent attention related to factors that
facilitate or impede the use of evidence-based interventions
in schools has revealed that teacher support and training are

primary reasons for the use of effective strategies (Bambara,
Nonnemacher, & Kern, 2009; Klingner, Arguelles, Hughes,
& Vaughn, 2001). The present findings extend previous
literature by providing descriptive support to the notion
that teacher procedural adherence is impacted by the level
of support and training provided. Furthermore, the results
of this study provide insight into the relation of social
validity, working alliance, and educator burnout. In terms
of social validity, analysis of the individual association
with measures of treatment integrity supported hypotheses
within the literature that it is related to greater treatment
integrity (e.g., Lentz et al., 1996). Such a finding should
offer greater insight into the relation of social validity and
treatment integrity within applied settings. In this regard,
the present study is an extension of previous literature
using direct observation measures of treatment integrity in
applied settings to study the relationship between social
validity and treatment integrity as recommended by All-
inder and Oats (1997). Although social validity was shown
to contribute to levels of treatment integrity in teachers,
there was a more powerful predictor included in the model:
working alliance.

For working alliance, it seems as though the coach–
teacher relationship might be an important factor to con-
sider regarding the procedural adherence of evidence based
practices in school-based settings. In fact, working alliance
was shown to impact treatment integrity above social
validity, which has been theorized as a prime factor in
treatment usage and integrity for more than two decades
(e.g., Witt & Elliott, 1985). It might be surprising to some
that integrity would be so heavily intertwined with teacher
perception of the coach–teacher relationship. However, the
importance of working alliance to treatment outcomes has
been previously identified in the field of psychotherapy
(e.g., Martin et al., 2000). Given the importance of treat-
ment fidelity to the magnitude of observed treatment

outcomes, measures of procedural fidelity are important variables to consider in designing classroom-based intervention packages. Therefore, identifying mechanisms to ensure increased usage and appropriate application are critical to the long-term sustainability of effective school practices. Working alliance might be one such mechanism to facilitate the integration of innovative classroom practices in school settings.

Finally, the impact of educator burnout was not shown to have a direct relationship with the level of treatment implementation. However, it was shown to moderate the effect of working alliance on treatment integrity for teachers reporting low levels of alliance. This finding is important for two reasons. First, it lends credence to the notion that teacher stress might impact implementation of evidence-based practices (Greenberg et al., 2001; Han & Weiss, 2005). According to the present study, this is particularly true for teachers that do not have a strong support system in place to assist with or encourage treatment implementation. For scaled-up interventions within applied settings, working alliance might be akin to administrative or co-worker support. Second, the observed interaction between educator burnout and working alliance provides additional evidence of the potential importance of the coach–teacher relationship. Specifically, strong working alliances were able to reduce the impact of educator burnout on treatment integrity. Teacher use and implementation of evidence-based practices might increase with the development of coaching procedures to assist with different stages of the validation process. In other words, good coaching practices from initial validation of a given practice through scaling up might lead to increased adoption of effective school practices.

Limitations

To interpret results of the present study, it is useful to consider its limitations. First, coaches were responsible for collecting treatment fidelity on each of their teachers, which could have results in biased assessment of treatment implementation. In addition, there were no interobserver agreement checks taken to ensure that observational data were reliable. This was necessitated by relatively limited project resources (i.e., time and money), but may have increased the likelihood of rater drift, coach reactivity, and the possibility of unreliable data. Second, although we focused on what we believe to be the most important coach factors, other coach variables might have provided further insight into the effects of the teacher–coach relationship. Examples of relevant coach variables might include certain personality, presentation, or demographic variables that would impact teacher procedural fidelity. Finally, in the present study fidelity measures were not collected from the

control teachers which meant that we were not able to directly assess the effects of the treatment condition on the teacher coach relationship.

Implications for Research and Practice

The present study has potential implications for addressing the research to practice gap that currently faces the field of school-based intervention research. As previously discussed, a primary concern for educational researchers is the willingness and ability of teachers and school-based personnel to adopt and implement evidence-based practices (Walker, 2004). Developing methods to ease the transition of promising practices from initial validation to scaled-up models should be a priority. As such, the school-based coach has been employed in many school districts to assist teachers implement research-based programs (Joyce & Showers, 1995). Fortunately, research has indicated that school coaches are generally effective for increasing the integrity of adopted interventions (Kretlow & Bartholomew, 2010). The findings of the present study might have helped to identify a mechanism that might ultimately promote or inhibit the effect of school coaches on teacher implementation. That is, the quality of the perceived relationship, or working alliance, between teacher and coach might be an important factor in the ability and willingness of teachers to use programs with integrity. It should be noted that these are only preliminary findings and additional research is needed to verify these conclusions. Further research should focus on ensuring that measures of working alliance are both valid and reliable; identifying coaching strategies that increase the likelihood of strategies being implemented with integrity; and determining those factors that might contribute to the development of strong or weak alliances. Although the present study does not provide a basis for making recommendations about effective coaching strategies, these research areas might assist with developing coaching procedures that would ultimately assist with ensuring that educational programs are delivered with integrity.

References

Allinder, R. M., & Oats, R. G. (1997). Effects of acceptability on teachers' implementation of curriculum-based measurement and student achievement in mathematics computation. *Remedial and Special Education, 18*(2), 113.

Bambara, L. M., Nonnemacher, S., & Kern, L. (2009). Sustaining school-based individualized positive behavior support: Perceived barriers and enablers. *Journal of Positive Behavior Interventions, 11*, 161–176.

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

Bierman, K. (2002). The implementation of the fast track program: An example of a large-scale prevention science efficacy trial. *Journal of Abnormal Child Psychology, 30*(1), 1–17.

Carnine, D. (1995). Trustworthiness, useability, and accessibility of educational research. *Journal of Behavioral Education, 5*(3), 251–258.

Cooley, E., & Yovanoff, P. (1996). Supporting professionals-at-risk: Evaluating interventions to reduce burnout and improve retention of special educators. *Exceptional Children, 62*(4), 336–355.

Davies, S., & Witte, R. (2000). Self-management and peer-monitoring within a group contingency to decrease uncontrolled verbalizations of children with attention-deficit/hyperactivity disorder. *Psychology in the Schools, 37*(2), 135–147.

Deshler, D. D. (2003). Intervention research and bridging the gap between research and practice. *Learning Disabilities: A Contemporary Journal, 1*(1), 1–7.

Dolan, L., Turkkan, J., Werthamer-Larsson, L., & Kellam, S. G. (1989). *The good behavior game training manual.* Baltimore, MD: Johns Hopkins Prevention Research Center.

Elliott, S. N. (1988). Acceptability of behavioral treatments: Review of variables that influence treatment selection. *Professional Psychology, Research and Practice, 19*(1), 68–80.

Elmore, R. F. (1996). Getting to scale with successful educational practices. *Harvard Educational Review, 66*, 1–27.

Embry, D. D. (2002). The good behavior game: A best practice candidate as a universal behavioral vaccine. *Clinical Child and Family Psychology Review, 5*(4), 273–297.

Evers, W., Brouwers, A., & Tomic, W. (2002). Burnout and self-efficacy: A study on teachers' beliefs when implementing an innovative educational system in the Netherlands. *British Journal of Educational Psychology. 72*, 227–243.

Fitzpatrick, M., & Knowlton, E. (2009). Bringing evidence-base self-directed intervention practices to the trenches for students with emotional and behavioral disorders. *Preventing School Failure, 53*, 253–264.

Fixen, D. L., Blasé, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation components. *Research on Social Work Practice, 19*, 531–540.

Greenberg, M. T., Domitrovich, C., & Bumbarger, B. (2001). The prevention of mental disorders in school-aged children: Current state of the field. *Prevention & Treatment, 4*(1), 1–52.

Gresham, F. M., & Lopez, M. F. (1996). Social validation: A unifying concept for school-based coaching research and practice. *School Psychology Quarterly, 11*(3), 204–227.

Guin, K. (2004). Chronic teacher turnover in urban elementary schools. *Education Policy Analysis Archives, 12*(42), 1–30.

Han, S. S., & Weiss, B. (2005). Sustainability of teacher implementation of school-based mental health programs. *Journal of Abnormal Child Psychology, 33*(6), 665–679.

Horvath, A. O., & Luborsky, L. (1993). The role of the therapeutic alliance in psychotherapy. *Journal of Consulting and Clinical Psychology, 61*(4), 561–573.

Horvath, A. O., & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology, 38*(2), 139–149.

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*, 88–110.

Johnson, M. R., Turner, P. F., & Konarski, E. A. (1978). The good behavior game: A systematic replication in two unruly transitional classrooms. *Education & Treatment of Children, 1*(3), 25–33.

Joyce, B., & Showers, B. (1995). *Student achievement through staff development: Fundamentals of school renewal.* New York, NY: Longman.

Kennedy, C. H. (2005). *Single-case designs for educational research.* Boston: Allyn and Bacon.

Klingner, J. K., Arguelles, M. E., Hughes, M. T., & Vaughn, S. (2001). Examining the schoolwide "spread" of research-based practices. *Learning Disability Quarterly, 24*, 221–234.

Kreft, I. G. G., & De Leeuw, J. (1998). *Introducing multilevel modeling.* Newbury Park: Sage Publications Ltd.

Kretlow, A. G., & Bartholomew, C. C. (2010). Using coaching to improve the fidelity of evidence-based practices: A review of studies. *Teacher Education and Special Education, 33*, 279–299.

Lentz, F. E., Allen, S. J., & Ehrhardt, K. E. (1996). The conceptual elements of strong interventions in school settings. *School Psychology Quarterly, 11*, 118–136.

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 68*(3), 438–450.

Maslach, C., Jackson, S. E., & Schwab, R. L. (1986). Educators survey. In C. Maslach & S. E. Jackson (Eds.), *Maslach burnout inventory manual.*

Medland, M. B., & Stachnik, T. J. (1972). Good-behavior game: A replication and systematic analysis. *Journal of Applied Behavior Analysis, 5*(1), 45.

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science & Practice, 12*(4), 365.

Ransford, C. R., Greenberg, M. T., Domitrovich, C. E., Small, M., & Jacobson, L. (2009). The role of teacher's psychological experience 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. (2002). *Hierarchical linear models: Applications and data analysis methods.* (2nd edn.) London: Sage.

Reimers, T. M., Wacker, D. P., & Koeppl, G. (1987). Acceptability of behavioral treatments: A review of the literature. *School Psychology Review, 16*, 212–227.

Schwab, R. L., Jackson, S. E., & Schuler, R. S. (1986). Educator burnout: Sources and consequences. *Educational Research Quarterly, 10*(3), 14–30.

Seeley, J. R., Small, J. W., Walker, H. M., Feil, E. G., Severson, H. H., Golly, A. M., et al. (2009). Efficacy of the first step to success intervention for students with attention-Deficit/Hyperactivity disorder. *School Mental Health, 1*(1), 37–48.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Singh, N. N., & Oswald, D. P. (2004). Evidence-based practice. Part I: General methodology. *Journal of Child and Family Studies, 13*(2), 129–142.

Snijders, T. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* Newbury Park: Sage Pubns Ltd.

Sterling-Turner, H. E., Watson, T. S., & Moore, J. W. (2002). The effects of direct training and treatment integrity on treatment outcomes in school coaching. *School Psychology Quarterly, 17*(1), 47–77.

Tingstrom, D. H., Sterling-Turner, H. E., & Wilczynski, S. M. (2006). The good behavior game: 1969–2002. *Behavior Modification, 30*(2), 225.

Walker, H. M. (2004). Commentary: Use of evidence-based interventions in schools: Where we've been, where we are, and where we need to go. *School Psychology Review, 33*(3), 398–408.

Witt, J. C., & Elliott, S. N. (1985). Acceptability of classroom intervention strategies. *Advances in School Psychology, 4*, 251–288.

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*(2), 203.

Yopp, D., Burroughs, E. A., Luebeck, J., Heidema, C., Mitchell, A., & Sutton, J. (2011). How to be a wise consumer of coaching strategies: Strategies teachers can use to maximize coaching's benefits. *Journal of Staff Development, 32*, 50–53.