Article

HAMMILL INSTITUTE ON DISABILITIES

The Journal of Special Education
2014, Vol. 48(3) 214–224

DOI: 10.1177/0022466913475872
journalofspecialeducation
.sagepub.com
SAGE

An Analysis of Preference Relative to Teacher Implementation of Intervention

LeAnne D. Johnson, PhD[1], Joseph H. Wehby, PhD[2], Frank J. Symons, PhD[1], Tara C. Moore, PhD[3], Daniel M. Maggin, PhD[4], and Kevin S. Sutherland, PhD[5]

Abstract

The purpose of this study was to conduct a preference trial as a preliminary test of preference effects on teacher behavior relative to implementation (adoption, adherence, quality). Teachers were randomly assigned to "preference" or "no-preference" groups and then trained to implement the intervention. Direct observation occurred immediately after initial training, after 6 weeks of coaching support, and after 4 weeks of no support. Results showed that, when compared with the no-preference group, teachers who had the opportunity to exert a preference adopted the intervention sooner and sustained higher fidelity and quality of implementation independent of coaching. Furthermore, though most teachers in the no-preference group did adopt the intervention and demonstrate high fidelity following coaching, implementation did not sustain after the withdrawal of coaching.

Keywords

intervention implementation, behavioral intervention, preference, coaching

There is a recognized need to improve the efficiency with which school-based interventions with established efficacy are translated into effective interventions and adopted for widespread use (i.e., brought to scale) in schools (Durlak & DuPre, 2008; Gresham, 2004; Han & Weiss, 2005; Smith, Daunic, & Taylor, 2007). The importance of efficiency and transportability of efficacious interventions is particularly urgent in our efforts to address the needs of students who engage in challenging behaviors that impede academic success (Gable, Hendrickson, & Van Acker, 2001; Malecki & Elliot, 2002; Smith et al., 2007; B. Walker, Cheney, Stage, & Blum, 2005). The widespread adoption of systemic approaches such as those based on a positive behavioral support framework (Sugai & Horner, 2008) has shown that despite a system's initial adoption of an evidence-based practice (EBP), implementation at the classroom level continues to be a major limitation in efforts to successfully scale up EBPs (Fagan, Hanson, Hawkins, & Arthur, 2008; Fairbanks, Simonsen, & Sugai, 2008; Greenwood, 2009; Kincaid, Childs, Blase, & Wallace, 2007; Lane, Bocian, MacMillan, & Gresham, 2004).

Several conceptual models have emerged that may be used to guide systematic examinations of variables hypothesized to influence the adoption and implementation of interventions. Recognizing the dynamic and nested effects across different systemic levels (external, organizational, intervention, and interventionist), Hagermoser Sanetti and Kratochwill (2009) provided a comprehensive summary of the variables relevant to adoption and implementation as identified by several research groups (Bosworth, Gingiss, Pothoff, & Roberts-Gray, 1999; Durlak & DuPre, 2008; Dusenbury, Brannigan, Falco, & Hansen, 2003; Fixsen, Naoom, Blase, Friedman, & Wallace, 2005; Gresham, 1989; Perepletchikova & Kazdin, 2005). Examples of relevant variables include (a) the level of support or opposition from stakeholders in the environment that is external to the organization adopting and implementing the intervention; (b) adequacy of funding, staffing, and leadership within the organization adopting and implementing the intervention; (c) the complexity and effort required to implement the intervention; and (d) the perceived need and buy-in of the interventionist related to the intervention being adopted and implemented. Building on the identification of these

[1]University of Minnesota, Minneapolis, USA
[2]Peabody College of Vanderbilt University, Nashville, TN, USA
[3]The University of Tennessee, Knoxville, USA
[4]University of Illinois at Chicago, USA
[5]Virginia Commonwealth University, Richmond, USA

Corresponding Author:
LeAnne D. Johnson, Department of Educational Psychology, University of Minnesota, 250 Education Science Building, 56 E. River Road, Minneapolis, MN 55455, USA.
Email: chaf0032@umn.edu

variables, others have attempted to conceptualize the processes, or pathways, through which different variables within and between the four levels described by Hagermoser Sanetti and Kratochwill interact to influence the initial willingness to adopt new practices as well as implement those practices over time (August, Gewirtz, & Realmuto, 2010; Elliott & Mihalic, 2004; Han & Weiss, 2005).

As understanding of implementation grows, particular attention is being given to scaling up implementation across a broad range of implementers and systems. Fixsen et al. (2005) described a process model that conceptualizes an entire system as moving through multiple stages of implementation from exploration through sustainability. It is in the early stages of exploration that factors known to affect fidelity of implementation are evaluated for contextual fit. Despite efforts to adopt EBPs that emerge from carefully planned explorations into how the EBP fits within an instructional context to meet the needs of the intended recipients, overall effectiveness continues to be dependent on the individual classroom teacher to make the choice to not only adopt the EBP, but also implement the EBP with high fidelity.

The decision to adopt an EBP may be conceptualized within the adaptive model of August et al. (2010). In an adaptive model, EBPs are selected in one of two ways, based on identified needs or preferences. Research in health-related fields provides evidence that provider and patient preferences about treatments are influential to subsequent treatment adherence in clinical trials (Bradley, 1993; Janevic et al., 2003; Ward et al., 2000; Wills & Holmes-Rovner, 2006). August et al. extend health-related research findings to preference-based models for enhancing parent implementation of interventions to address children's social-emotional needs. Knowing the complexity of obtaining information to make a priori decisions to adapt EBPs to fit individual family contexts, August et al. and others (Hoza, Johnson, Pillow, & Ascough, 2006; Morrissey-Kane & Prinz, 1999; Nock, Ferriter, & Holmberg, 2007) assert that the simple opportunity to express a preference about an intervention may allow the implementer to make a decision based on the salient influences at the time. The adaptive model of August et al. holds promise in furthering efforts to scale up implementation of EBPs by treating the preference of the interventionist from the onset as a critical variable. In education, despite theoretical models that highlight the importance of implementer buy-in and perceptions about an EBP, there has not yet been an empirical examination of how the simple opportunity to allow teachers to exert a preference relates to actual adoption and implementation.

After the initial decision is made by an individual interventionist to adopt an EBP, the implicit hope is that a commitment to high-fidelity implementation is part of that decision. Years of research examining fidelity of implementation suggests that though the commitment to adopting a practice is necessary, it is not sufficient for ensuring fidelity (Durlak & DuPre, 2008; Fagan et al., 2008). Several models exist for providing partnership-based consultation and coaching (Kelleher, Riley-Tillman, & Power, 2008; Power et al., 2005; Sheridan, Swanger-Gagne, Welch, Kwon, & Garbacz, 2009; Stormont & Reinke, 2012) that support coaching as an effective means for enhancing fidelity of implementation at the classroom level (Barrett, Bradshaw, & Lewis-Palmer, 2008). Within collaborative coaching models, the assumption is that highly skilled coaches are able to guide the implementation of any given EBP within the individualized contextual needs of the classroom and teacher. What remains unclear is how coaching may interact with the way in which a teacher came to be implementing an intervention. In other words, what relation does coaching have to fidelity when a teacher entered into the implementation process through either having the opportunity to exert a preference for the selected EBP or simply being told which EBP they will be coached to implement? It is at the level of the individual interventionist that examinations of the process associated with adoption and implementation need to be examined if EBPs are to truly elevate the effectiveness of the field (Greenwood, 2009).

To date, despite discussions of the influence of preferences on implementation, there has not yet been a randomized controlled preference trial conducted with teachers choosing among EBPs in classroom settings. The purpose of this study was to conduct a preference trial to study implementation effects (number of adopters and fidelity) across a sample of teachers as a "proof of concept" test for the relevance of preference on teacher decision making and implementation.

Method

Participants and Settings

At the onset, a total of 69 teachers (88% female; 68% general education, 32% special education) working with kindergarten through sixth-grade students participated in this study. Participants were identified from 14 schools distributed across three large metropolitan school districts from three regions of the United States. Districts ranged in size from 23,200 to 70,140 enrolled students with an average racial demographic of 24% White, 58% Black, 12% Hispanic, 4% Asian, and 2% American Indian. Across all three districts, an average of 12% of students had limited English proficiency, 15% received special education services, and 69% qualified for free or reduced lunch. To recruit teachers, district or school administrators distributed information to invite their attendance at an informational meeting about this study. During the meeting, teachers were given a brief overview of the intervention strategies and told that if they chose to participate in the study, they

would receive training and support for implementation of one of the interventions. Participating teachers averaged 12 years of teaching experience ($SD = 9.5$) with 44% having obtained master's degrees or higher, 33% of teachers were Black, 58% White, 2% Hispanic, and 7% did not report their racial background.

Assignment to Preference Versus No-Preference Groups for Implementation and Analysis

Each of the 69 participating teachers was randomly assigned to one of the three groups: "preference group" (P; $n = 25$), "no preference–good behavior game" (NP-GBG; $n = 21$), or "no preference–teacher self-monitoring" (NP-TSM; $n = 23$). Random assignment by teacher was conducted within blocks for each of the three school districts using an online randomization program (www.graphpad.com). Teachers randomly assigned to the "preference group" met individually with a project consultant who described the procedures for both the GBG and TSM (see below for more information). The GBG and TSM were selected for this project given the emerging, strong, evidence base for their use as strategies that modify the classroom learning environment through decreases in students' challenging behaviors that result from changes in teachers' behaviors (Epstein, Atkins, Cullinan, Kutash, & Weaver, 2008).

The procedures for both GBG and TSM were manualized and presented to each teacher such that both interventions appeared relatively equivalent in terms of time and effort required of the teacher. More specifically, if implementing GBG, teachers were asked to follow the original procedures for GBG described by Barrish, Saunders, and Wolf (1969) and plan a 10-min portion of their language arts instruction in which they would implement the group-oriented behavior management contingency. The expectation was set that teachers implement GBG for 10 min, 5 days per week. If implementing TSM, teachers were asked to follow procedures described by Sutherland, Wehby, and Copeland (2000) and again plan a 10-min portion of their language arts instruction in which they would wear the tape recorder and record their instruction on 3 days each week. To maintain perceptions of equivalent time and effort compared with GBG, 3 days of recording was selected to offset time requirements for TSM, which also involved teachers listening to a 5-min portion of each audiotape and recording their behaviors at the end of the week.

Once procedures were described and clarified for individual teachers, each teacher that had been randomly assigned to the preference group was then asked to make a choice between implementing GBG or TSM. Twenty-two (88%) teachers randomly assigned to the "preference group" selected the GBG to implement. Consequently, given the overwhelming preference for GBG, all of the following procedural descriptions and subsequent analyses of the effect of preference on implementation are based only on the 43 participants who implemented GBG. More specifically, the two comparison groups for all subsequent descriptions of this study are both implementing GBG and distinguished only by those teachers who were randomly assigned to express a preference for implementing the procedure (P-GBG; $n = 22$) or were assigned to implement it with no opportunity to indicate a preference (NP-GBG; $n = 21$).

Measures

Direct observation of implementation. As a preference trial designed to explicitly examine the relation between teacher preferences and teacher implementation, the primary dependent variables in this study are related to implementation. Measures of child behavior and classroom climate were ancillary to the primary aims and therefore not gathered. Direct observation of teacher's implementation was monitored at three preassigned observation intervals—initial week of implementation, immediately post 6 weeks of coaching, and at a 4-week follow-up. During 6 weeks of coaching, observations of implementation of the GBG occurred each week to guide feedback and support. Following 6 weeks of coaching, there was no contact with teachers until the 4-week follow-up observation. All observations were scheduled with the teacher prior to conducting the observation. During each observation, implementation was evaluated using three different variables: (a) number of actual implementers, (b) percentage of procedural items, and (c) quality of implementation. Observers at each time point were blind to the group assignment of teacher implementers.

Number of actual implementers. At each observation, observers recorded whether the intervention was actually being implemented by the teacher. Teachers were only coded as an actual implementer if implementation was directly observed by the project observer. Therefore, all observations were prearranged with all teachers. For analysis, the number of teachers who were actually implementing the intervention was summed to provide a total number of actual implementers at each of the three predetermined observation intervals.

Percentage of procedural items. Teachers were trained to implement the intervention procedure in direct adherence to the procedures listed on the fidelity checklist. The GBG fidelity checklist consists of 14 procedural items: (a) announcing game before beginning, (b) referring to teams before beginning, (c) referring to classroom rules, (d) referring to requirement to win, (e) referring to rule violation process, (f) indicating the start of the game, (g) responding immediately

to rule violations, (h) responding consistently to every rule violation, (i) responding with a typical instructional voice, (j) identifying the child/team violating rule, (k) praising children or teams, (l) reviewing scores, (m) handing out prizes or delivering reinforcers, and (n) playing game with a clear beginning, middle, and end. To establish percentage of procedural items implemented, each item was dichotomously coded as observed or not observed. A column was also included for documenting whether there was other evidence for completion of a procedure not directly observed. For example, if "handing out of prizes or delivering reinforcers" involved the teacher telling students that they would receive 5 min of free time at the end of that instructional period, that was recorded as other evidence for implementation of that procedure given that the consultant likely would not be present to observe the actual free time. For analysis, these columns were collapsed to obtain an overall percentage of procedural items implemented.

Four procedural items (Items g, h, i, and j) could only be coded if disruptive behavior occurred while the intervention was being implemented. Of the 126 observations that were analyzed for this study, seven (distributed across six different teachers at different time points) included implementation when no disruptive behaviors were observed. If no disruptive behavior was observed, those items were not coded and the total number of items used to calculate the percentage of procedural items implemented for the observation was reduced to 10. This approach created a more conservative estimate of fidelity that also more accurately represented the context of implementation for that very small subset of observations.

Quality of implementation. For each procedural item observed, a quality of implementation score was also assigned to quantify variability in the degree to which individual teachers implemented each item. For example, one procedural item states, "refer to requirements to win." For this item, if the teacher made any reference to how teams win the game, the dichotomous coding procedural items implemented would reflect that that particular item was implemented. However, there were quantitative and qualitative differences when a teacher simply stated, "you all remember how to win" versus stating, "when the timer goes off, if your team has less than three checks on the board, your whole team will win for the day and your names go up on the leader board." To quantify this variability in implementation, each procedural item was assigned one of five possible scores on a scale of 0% to 100%; 0% represented that the item was not implemented at all, 25% represented minimal quality of implementation with significant room for improvement, 50% represented half/partial quality with some room for improvement, 75% represented good quality with only minimal room for improvement, and a score of 100% represented that the item was implemented with the highest possible quality. For analysis, individual teacher's quality scores across all procedural items were summed and then divided by the total number of items to obtain a mean quality of implementation score for each observation.

Reliability and internal consistency of the observation system. The observation system was initially developed and tested across three sites within a larger intervention study conducted during the academic year prior to the study described here (Wehby, Maggin, Moore Partin, & Robertson, 2012). In that study, a total of 10 observers completed a total of 607 observations distributed across 51 teacher participants. Of the 10 observers, 8 continued to conduct observations for the present study. The reliability of the observation system in this study was established by calculating intraclass correlation coefficients (ICCs) in which two observers simultaneously completed nine observations. ICC for the percentage of procedural items checklist was .97 (95% confidence interval [CI] = [0.95, 0.99]), for quality of implementation ratings was .95 (95% CI = [0.91, 0.98]), and for recordings of actual implementation was 1.0 (95% CI = [1.0, 1.0]). For the checklist reported in this study, analysis of the internal consistency produced an alpha coefficient of .97.

Procedure

Training and coaching to support implementation of GBG. The GBG (Barrish et al., 1969; Kellam, Ling, Merisca, Brown, & Ialongo, 1998) is a group-contingency classroom management procedure designed to reduce problem behavior in the classroom. Research has documented the effectiveness of the GBG in decreasing levels of aggression and disruption as well as increasing on-task behavior during instructional times (Dolan et al., 1993; Harris & Sherman, 1973; Ialongo et al., 1999; Kellam et al., 1998; Medland & Stacknik, 1972). The GBG is based on implementation of explicit rules as well as explicit and consistent teacher responses to students' rule-following and rule-violating behaviors. For this study, teachers were initially trained in the manualized procedures for GBG by a project coach during a 45- to 60-min meeting. Each coach reviewed the manual that described the procedural steps for implementation (in checklist format and a weekly implementation guide) and identified the 10-min duration that GBG would be implemented within his or her typical classroom routines involving language arts instruction. All teachers were trained to implement the intervention for a minimum of 10 min per day for 5 days per week. Following the initial training, all teachers (P-GBG and NP-GBG) received weekly coaching from a project coach for 6 weeks. Teachers were explicitly told that behavior coaches would be observing each week and available to support their implementation for only the first 6 weeks. After the 6-week period of coaching, teachers

in both groups were told that observers would be back after 4 weeks to document if and how GBG was still being implemented.

Behavioral coaches were graduate-level university students with at least 1 year experience providing feedback to teachers implementing the GBG within systematic research activities. Given that the GBG was a component of a treatment package within a larger parent project at the time, all behavior coaches had at least 1 year of previous experience observing and giving feedback about the manualized procedures for implementing GBG. The weekly support from the behavioral coach included any combination of the following after the weekly observation of the teacher's implementation had occurred: (a) providing the resources needed to initiate implementation such as a timer, posters identifying the rules and teams, as well as a system for visually tracking rule violations while the game is being played; (b) modeling implementation for the teacher during instructional sessions; and (c) providing explicit feedback based on the procedural checklist used to monitor fidelity of implementation.

Data Analyses

Repeated-measures ANOVA (RM ANOVA) was used to examine overall differences between the P-GBG and NP-GBG groups on percentage of procedural items implemented and implementation quality across the three observation intervals. Planned post hoc ANOVA was conducted at each time point to further examine differences between the groups. Chi-square tests were conducted for each observation interval to test for any differences in the number of teachers actually implementing the intervention. All statistical analyses were conducted using Version 19 of the SPSS Statistics Software.

Results

Data gathered across each time point (initial exposure, post 6-week coaching, and 4-week follow-up) were screened for missing values, normality, and homogeneity of variance, within each time point. Screening of the data prior to the analysis confirmed that implementation data were missing at the third observation interval for one case due to a maternity leave. This case was excluded from subsequent analyses that resulted in the P-GBG and NP-GBG groups now each having 21 participants. Figure 1 provides a visual display of the mean performance of both groups on each of the three implementation variables over time.

Number of Actual Implementers

Examining the relation between the opportunity to exert preference and number of implementers involved two complimentary approaches. First, differences in the numbers of implementers at each of the three time points were evaluated using a series of chi-square comparisons between the P-GBG and NP-GBG groups (see Figure 1, top panel). Significant differences were observed between groups at the initial observation session ($\chi^2 = 5.32$, $p = .02$, $\Phi = .35$) with 77% of the teachers in the P-GBG group implementing compared with 43% of the NP-GBG group. No significant

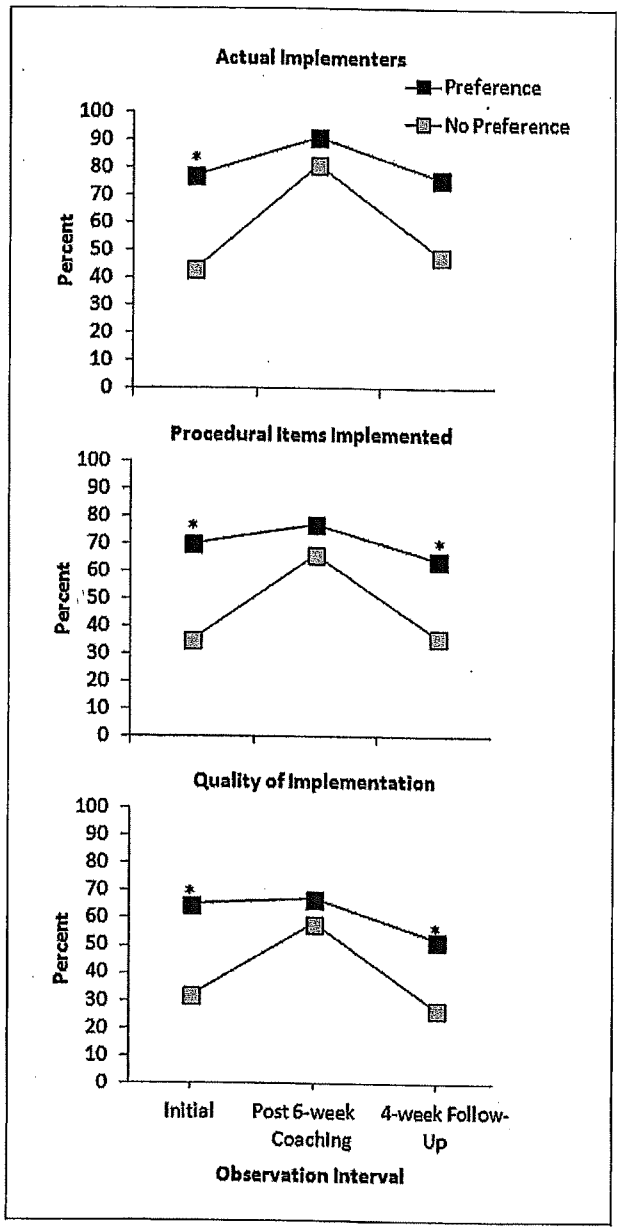

Figure 1. Mean percentage of actual implementers (top panel), procedural items implemented (middle panel), and quality of implementation (bottom panel) by group across three observation intervals.
*Statistically significant difference between groups.

Table 1. Descriptive Summary of Implementation by Subgroups Identified by Point of Adoption.

	Preference ($n = 21$)			No preference ($n = 21$)		
Implementation variables	Adopt at initial	Adopt post-coaching	Never adopt	Adopt at initial	Adopt post-coaching	Never adopt
Number of actual implementers at point of adoption	16	3	2	9	9	3
Number of actual implementers sustaining implementation beyond point of adoption	14	2	—	8	2	—
Mean percentage of procedural items implemented at adoption	90% (15%)	81% (18%)	—	83% (13%)	75% (31%)	—
Mean percentage of procedural items implemented beyond point of adoption	84% (23%)	81% (18%)	—	77% (26%)	18% (37%)	—
Mean quality of implementation at adoption	84% (18%)	55% (7%)	—	75% (15%)	66% (24%)	—
Mean quality of implementation beyond point of adoption	69% (28%)	35% (34%)	—	60% (27%)	16% (34%)	—

Note. Numbers in parentheses represent standard deviations.

difference in the number of implementers between groups was found at the post 6-week consultation observation ($\chi^2 = .89, p = .31, \Phi = .14$) with 91% of the P-GBG group implementing and 81% of the NP-GBG group implementing. At the final follow-up observation conducted 4 weeks later, there was a marginally statistically significant difference between groups demonstrating actual implementation at proportions similar to the initial observation (P-GBG = 76%, NP-GBG = 48%; $\chi^2 = 3.64, p = .06, \Phi = .29$).

To understand implementation at the level of the individual classroom teacher, teachers from both the P-GBG and NP-GBG groups were further classified. As displayed in Table 1, teachers were categorized by either demonstrating adoption of the intervention, through observed implementation, at the initial observation session, post 6 weeks of coaching, or as never demonstrating adoption of the intervention. In both groups, there were small numbers of teachers who never demonstrated adoption of GBG (P-GBG = 2 or 10%, NP-GBG = 3 or 14%). All other teachers either adopted GBG immediately (P-GBG = 16 or 76%, NP-GBG = 9 or 43%) or adopted it by the time coaching was complete (P-GBG = 3 or 14%, NP-GBG = 12 or 51%). These findings are consistent with those displayed in Figure 1 in which a significant difference in the number of implementers was observed at the initial observation, but that no difference was observed immediately following coaching. More importantly, of those teachers who adopted GBG immediately, most sustained implementation through follow-up, though teachers in the preference group outnumbered the no-preference group nearly 2:1. Furthermore, though the two groups were comparable in number of actual implementers following the 6 weeks of coaching, Table 1 also highlights that despite the increase in the number of teachers demonstrating adoption of GBG in the no-preference group after coaching, 78% of those teachers (seven of the nine) did not sustain implementation to the 4-week follow-up observation.

Percentage of Procedural Items

Beyond identifying patterns of how many teachers were implementing the intervention, fidelity of implementation was examined using RM ANOVA to evaluate within- and between-group differences in percentage of procedural items implemented across time (see Figure 1, middle panel). Distributions for both the P-GBG and NP-GBG groups met the assumptions of normality with nominal negative skews for the P-GBG group (range = −0.81 to −1.9) and homogeneity of variance over time based on Levene's test, $F(1, 40)$ range = 1.33 to 2.58, p range = .12 to .26. The omnibus test resulted in statistically significant main effects for group membership (P-GBG/NP-GBG), $F(1, 40) = 5.56, p = .02$, partial $\eta^2 = .12$ with observed power of .63, and time, $F(2, 39) = 6.3, p < .01$, partial $\eta^2 = .24$ with observed power of .87. A statistically significant interaction between group and time was not observed, $F(2, 80) = 1.43, p = .25$, partial $\eta^2 = .03$. Planned comparisons of the differences in the estimated means between groups at each time point were conducted using three univariate ANOVAs. At the initial observation, the P-GBG group implemented a significantly higher percentage of the procedural items compared with the NP-GBG group, P-GBG: $X = 70\%$, $SD = 41\%$; NP-GBG: $X = 35\%$, $SD = 43\%$; $F(1, 41) = 7.31, p = .01, d = .83$. There were no significant differences immediately following the 6-week consultation, P-GBG: $X = 77\%$, $SD = 28\%$; NP-GBG: $X = 66\%$, $SD = 39\%$; $F(1, 42) = 1.23, p = .27, d = .33$. At the final follow-up, the P-GBG group again demonstrated significantly higher levels of fidelity than the NP-GBG group, P-GBG: $X = 64\%$, $SD = 40\%$; NP-GBG: $X = 36\%$, $SD = 42\%$; $F(1, 39) = 4.85, p = .03, d = .68$.

Implementation of procedural components was also examined at the level of the individual teacher based on the previously described classification into groups representing point of adoption. Descriptively, though the number of

implementers differed, teachers who implemented at the initial observation did so at levels that appear comparable between the preference and no-preference groups (see Table 1). Similar levels of fidelity did sustain across the two remaining time points for teachers who adopted GBG at the initial observation. In contrast, for the subgroup of teachers who did not demonstrate adoption of GBG until after 6 weeks of coaching, slightly lower and more varied levels of fidelity were observed. Levels of fidelity for this subgroup of teachers did not maintain at similar levels over time. At follow-up, the preference group implemented at a higher yet quite varied level and the no-preference group implemented at a lower and varied level. Lower levels of fidelity and wide variability for each group are indicative of the number of teachers who demonstrated adoption of GBG following coaching, but were no longer implementing at follow-up.

Quality of Implementation

Quality of implementation was also examined as an important aspect of fidelity. RM ANOVA was used to examine within- and between-group differences in implementation quality across time (see Figure 1, bottom panel). The distributions for both groups met the assumptions of normality and homogeneity of variance based on Levene's test, $F(1, 40)$ range = .007 to 1.132, p range = .29 to .94. The omnibus test results were similar to those examining level of fidelity, with statistically significant main effects for group membership (P-GBG/NP-GBG), $F(1, 40) = 5.46$, $p = .02$, partial $\eta^2 = .12$ with observed power of .63, and time, $F(2, 39) = 8.69$, $p < .01$, partial $\eta^2 = .31$ with observed power of .96. A statistically significant interaction between group and time was not observed, $F(2, 38) = 1.95$, $p = .17$, partial $\eta^2 = .09$. Planned comparisons of the differences in the estimated means between groups at each time point were again conducted using three univariate ANOVAs. Consistent with findings for the percentage of procedural items implemented, the P-GBG group implemented with a higher quality of fidelity at the initial observation, P-GBG: $X = 65\%$, $SD = 39\%$; NP-GBG: $X = 32\%$, $SD = 39\%$; $F(1, 41) = 7.68$, $p \leq .01$, $d = .85$. There were no significant differences immediately following the 6-week consultation, P-GBG: $X = 67\%$, $SD = 28\%$; NP-GBG: $X = 58\%$, $SD = 34\%$; $F(1, 42) = .92$, $p = .34$, $d = .82$. At the final 4-week follow-up, teachers in the P-GBG group were again implementing with a significantly higher quality of fidelity, P-GBG: $X = 52\%$, $SD = 36\%$; NP-GBG: $X = 27\%$, $SD = 33\%$; $F(1, 39) = 5.28$, $p = .03$, $d = .53$.

Quality of implementation was further examined at the level of the individual teacher based on the previously described classification into subgroups (see Table 1). Consistent with implementation of procedural items, quality of implementation for the subgroup of teachers who adopted GBG from the onset appeared comparable between the preference and no-preference groups at the initial observation. Quality of implementation decreased for both groups over time and became more variable. Also consistent with implementation of procedural items, the subgroup of teachers who adopted following coaching implemented at overall lower levels of quality across both groups that decreased even further at follow-up.

Discussion

Analogous to the medical field in which patient adherence to a doctor's treatment recommendations is a key variable to evaluating changes in health status, teachers' fidelity of implementation is a key variable in education when evaluating changes in instructional behaviors that set the context for student learning. The purpose of this study was to experimentally examine implementation in relation to teacher preference of interventions designed to enhance classroom management of student behavior. Overall, the results, although preliminary, suggest that when given an opportunity to express a preference among at least two interventions, a greater number of teachers will implement the self-selected intervention with improved fidelity to the procedures and higher quality over time when compared with teachers who were not given the opportunity to express a preference. Furthermore, though coaching increased the number of implementers in the group that was not given the opportunity to express a preference, only a very small proportion of those teachers sustained implementation through follow-up. This finding indicates that simply having the opportunity to express a preference from the onset may yield higher levels of fidelity with a greater number of implementers, as evidenced by differences at the initial observation period that extended to over time.

Empirical Implications

Although we originally planned to evaluate across two specific interventions (GBG and TSM), because of the almost exclusive self-selection of GBG, the initial scope of the study was reduced to comparing teachers' self-selecting GBG compared with teachers who were assigned GBG. Thus, our results do not provide any evidence one way or the other about GBG compared with TSM. However, the fact that teachers so heavily favored GBG when given a choice suggests that some treatment-implementer factors (e.g., implementer preference) are likely operating in any multicomponent study, even if not measured, and therefore likely contributing to implementation variance. Several studies have conducted post hoc analyses to identify and classify implementer characteristics that explain variability in implementation (Evers, Brouwers, & Tornic, 2002; Kealey, Peterson, Gaul, & Dihn, 2000; Reimers, Wacker, &

Koeppl, 1987; Stein & Wang, 1988). Studies such as these have provided empirical evidence to support attention to an array of variables during stages of exploration and initial adoption on an EBP (Fixsen et al., 2005) such that sustainability at a systems level may be achieved through a good contextual fit. Achieving contextual fit at the classroom level is often supported through coaching to not only increase numbers of implementers, but also fidelity of implementation. Questions remain, however, for how to increase initial adoption of EBPs by individual implementers from the onset of a systems-level decision to adopt an EBP.

Participant preference-driven intervention trials are being proposed as enhancements to the standard randomized control trial design (see Janevic et al., 2003; Wills & Holmes-Rovner, 2006, for discussion). Preference-driven trials may provide an opportunity to systematically explore process variables that may affect the diffusion of EBPs on a larger scale. Adaptive interventions and research designs proposed by Collins, Murphy, and Bierman (2004) as well as August et al. (2010) highlight (a) identification of core features of an EBP that are hypothesized to be important to facilitating change, (b) understanding of contextual features that may affect fidelity of implementation, and (c) identification of a framework for considering how the actual EBP or the process of inducting the EBP into the system may be systematically adapted to tailoring variables based on the needs and/or preferences of the participants. Findings from this study suggest that adapting how researchers structure the manner by which interventionists approach implementation of an intervention (express preference or receive directive) may have a stronger influence than other means by which implementation is influenced (i.e., coaching). This study was designed to focus exclusively on that issue by taking an EBP with a well-documented empirical base of efficacy and simply manipulating the manner by which teachers approached adoption of the practice. As empirical evidence continues to establish the efficacy of specific interventions for specific students in specific instructional contexts, a parallel based of empirical evidence is needed to establish why and how implementation does or does not occur. This preliminary study was designed to initiate empirical examinations of implementation from the perspective of the implementer rather than the interaction between implementation and student outcomes.

Educational Implications

With efforts to identify necessary supports for implementation of EBPs in schools, it is particularly important to note that these findings are consistent with assertions that training with implementation feedback may enhance fidelity of implementation (Adelman & Taylor, 2003; Elliott & Mihalic, 2004; Han & Weiss, 2005). This is supported by the increase in both (fidelity and quality) variables from the initial to the post 6-week coaching observations as well as the marginal differences between the preference and no-preference groups at that post 6-week coaching time point. With examination of implementation at the level of the individual teacher, it became clear, however, that different patterns of implementation emerged relative to group assignment. Categorization of individual teachers within the preference and no-preference groups to subgroups was based on Rogers's (1995) conceptualization of the process through which individuals adopt innovations. Both groups in this study included "early adopters" who demonstrated high-fidelity implementation from the onset that maintained over time independent of coaching. However, though no causal relation is being suggested, the preference group was characterized by nearly twice as many early adopters than the no-preference group. The no-preference group included a higher proportion of "late adopters" who demonstrated adoption with high-fidelity implementation only after coaching had been provided.

Therefore, when circumstances are such that teachers are not given the opportunity to self-select interventions, offering regular and individualized coaching support may occasion improved fidelity of implementation. Unfortunately, higher levels of fidelity that are obtained under these circumstances may not maintain over time. The results from the present study demonstrate that teacher preference, expressed through self-selection of the intervention, may make a unique contribution to understanding variability in implementation. Future research is needed to examine the impact of a priori decisions to adapt either the actual EBP or the process by which the EBP is disseminated for adoption based on other tailoring variables (i.e., participant functioning, social-emotional needs, perceived benefit, outcome expectations, intensity/severity of targeted behaviors). Enhanced attention to the process of adaption and adoption of EBPs may provide resource efficient means to produce greater implementation in schools in light of limited capacity for regular, expert coaching.

Limitations and Future Directions

There are several study limitations that should be mentioned. Although there was random assignment, the (relatively small) sample was not randomly selected. Thus, the inferences made about the results should be limited most directly to the teachers and intervention involved. As mentioned, there was no opportunity to evaluate teacher preference in relation to "teacher self-monitoring," thus we have no basis one way or the other to make comparative statements about TSM and GBG other than to point out that the vast majority of teachers, when given a choice, selected GBG. This study did limit the number of treatment options available to the preference group to two (TSM and GBG). How and in what direction the opportunity to express a preference among a larger array of interventions (three or

more) would affect implementation is unknown and worth exploring.

The primary dependent variables for this study were all direct observations of teacher implementation as a more proximal outcome related to teacher preferences. Therefore, more distal outcome measures such as classroom atmosphere or student academic and behavioral performance were not included. Future trials that incorporate child outcomes relative to teacher preferences for implementation of interventions are necessary. Finally, this study did not include additional measures to complement the direct observational measures of teacher behavior. Future work adopting a "mixed methods" approach could be extremely important in documenting what happened (direct observation) with why it happened (interview). Using mixed methods within preference trials such as this may offer valuable opportunities to examine a variety of moderating variables such that empirically derived decision guides for selecting and implementing EBPs may be developed to support efforts to go to scale.

As noted by H. M. Walker (2004) and reviewed by Mihalic and Irwin (2003), the tendency to emphasize outcomes has come at the expense of understanding the process through which the outcomes are obtained. Only a very small percentage of studies report directly on implementation, and fewer still are the studies designed to examine it experimentally. H. M. Walker (2004) suggested, "Perhaps the greatest opportunity for improving understanding of applied interventions lies in the systematic study of the implementation process" (p. 403). The study reported here was a first step in that direction. Future work should continue to explore the range of variables affecting teacher preferences for one intervention over another and whether those same variables similarly affect the quality of implementation and student outcomes. Designing studies to do so may be more difficult than "first-order" efficacy studies, but the promise of reducing the research to practice gap may more fully be realized the sooner we undertake the challenge.

Authors' Note

The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant H324P040013 to Vanderbilt University.

References

Adelman, H. S., & Taylor, L. (2003). On sustainability of project innovations as systemic change. *Journal of Educational and Psychological Consultation, 14*, 1–26.

August, G. J., Gewirtz, A., & Realmuto, G. M. (2010). Moving the field of prevention science to service: Integrating evidence-based preventative interventions into community practice through adapted and adaptive models. *Applied & Preventative Psychology, 14*, 72–85.

Barrett, S. B., Bradshaw, C. P., & Lewis-Palmer, T. (2008). Maryland statewide PBIS initiative: Systems, evaluation, and next steps. *Journal of Positive Behavior Intervention, 10*, 105–114.

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*, 119–124.

Bosworth, K., Gingiss, P., Pothoff, S. J., & Roberts-Gray, C. A. (1999). A Bayesian model to predict the success of the implementation of health and education innovations in school centered programs. *Evaluation and Program Planning, 22*, 1–11.

Bradley, C. (1993). Designing medical and educational intervention studies. *Diabetes Care, 16*, 509–518.

Collins, L. M., Murphy, S. A., & Bierman, K. L. (2004). A conceptual framework for adaptive preventative interventions. *Prevention Science, 5*, 185–196.

Dolan, L. J., Kellam, S. G., Brown, H. C., Werthamer-Larsson, L., Rebok, G. W., Mayer, L. S., & Wheeler, L. (1993). The short-term impact of two classroom-based preventive interventions on aggressive and shy behaviors and poor achievement. *Journal of Applied Developmental Psychology, 14*, 317–345.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–350.

Dusenbury, L., Brannigan, 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.

Elliott, D., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science, 5*, 47–53.

Epstein, M., Atkins, M., Cullinan, D., Kutash, K., & Weaver, R. (2008). *Reducing behavior problems in the elementary school classroom: A practice guide* (NCEE #2008-012). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides

Evers, W. J., Brouwers, A., & Tornic, W. (2002). Burnout and self-efficacy: A student on teachers' beliefs when implementing and innovative educational system in the Netherlands. *British Journal of Educational Psychology, 72*, 227–244.

Fagan, A. A., Hanson, K., Hawkins, J. D., & Arthur, M. W. (2008). Bridging science to practice: Achieving prevention program implementation fidelity in the community youth development study. *American Journal of Community Psychology, 41*, 235–249.

Fairbanks, S., Simonsen, B., & Sugai, G. (2008). Classwide secondary and tertiary tier practices and systems. *TEACHING Exceptional Children, 40*(6), 44–52.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI Publication #231). Tampa: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network.

Gable, R. A., Hendrickson, J. M., & Van Acker, R. (2001). Maintaining integrity of FBA-based interventions in schools. *Education & Treatment of Children, 24*, 248–260.

Greenwood, C. R. (2009). Treatment integrity: Revisiting some big ideas. *School Psychology Review, 38*, 547–553.

Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review, 18*, 37–50.

Gresham, F. M. (2004). Current status and future directions of school-based behavioral interventions. *School Psychology Review, 33*, 326–334.

Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38*, 445–459.

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

Harris, W. V., & Sherman, J. A. (1973). Use and analysis of the "good behavior game" to reduce disruptive classroom behavior. *Journal of Applied Behavior Analysis, 6*, 405–417.

Hoza, B., Johnston, C., Pillow, D., & Ascough, J. C. (2006). Predicting treatment response for childhood attention-deficit hyperactivity disorders: Introduction of a heuristic model to guide research. *Applied and Preventive Psychology: Current Scientific Perspectives, 11*, 215–229.

Ialongo, N. S., Werthamer, L., Kellam, S. G., Brown, S. H., Wang, S., & Lin, Y. (1999). Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression, and antisocial behavior. *American Journal of Community Psychology, 27*, 599–641.

Janevic, M. R., Janz, N. K., Dodge, J. A., Lin, X., Pan, W., Sinco, B. R., & Clark, N. M. (2003). The role of choice in health education intervention: A review and case study. *Social Science & Medicine, 56*, 1581–1594.

Kealey, K. A., Peterson, A., Gaul, M. A., & Dihn, K. T. (2000). Teacher training as a behavior change process: Principles and results from a longitudinal study. *Health Education & Behavior, 27*, 64–81.

Kellam, S., Ling, X., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology, 10*, 165–185.

Kelleher, C., Riley-Tillman, T. C., & Power, T. J. (2008). An initial comparison of collaborative and expert-driven consultation on treatment integrity. *Journal of Educational and Psychological Consultation, 18*, 294–324.

Kincaid, D., Childs, K., Blase, K. A., & Wallace, F. (2007). Identifying barriers and facilitators in implementing schoolwide positive behavior support. *Journal of Positive Behavior Interventions, 9*, 174–184.

Lane, K. L., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2004). Treatment integrity: An essential-but often forgotten-component of school-based interventions. *Preventing School failure, 48*, 36–43.

Malecki, C. K., & Elliot, S. N. (2002). Children's social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly, 17*, 1–23.

Medland, M. B., & Stacknik, T. J. (1972). Good behavior game: A replication and systematic analysis. *Journal of Applied Behavior Analysis, 5*, 45–51.

Mihalic, S., & Irwin, K. (2003). Blueprints for violence prevention: From research to real-world settings—Factors influencing the successful replication of model programs. *Youth Violence and Juvenile Justice, 1*, 1–23.

Morrissey-Kane, E., & Prinz, R. J. (1999). Engagement in child and adolescent treatment: The role of parental cognitions. *Clinical Child and Family Review, 2*, 183–198.

Nock, M. K., Ferriter, C., & Holmberg, E. (2007). Parent beliefs about treatment credibility and expectancies for improvement: Assessment and relation to treatment adherence. *Journal of Child and Family Studies, 16*, 27–38.

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice, 12*, 365–383.

Power, T. J., Blom-Hoffman, J., Clarke, A. T., Riley-Tillman, T. C., Kelleher, C., & Manz, P. H. (2005). Reconceptualizing intervention integrity: A partnership-based framework for linking research with practice. *Psychology in the Schools, 42*, 495–507.

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

Rogers, E. (1995). *Diffusion of innovations* (4th ed.). New York, NY: Free Press.

Sheridan, S. M., Swanger-Gagne, M., Welch, G. W., Kwon, K., & Garbacz, S. A. (2009). Fidelity measurement in consultation: Psychometric issues and preliminary examination. *School Psychology Review, 38*, 476–495.

Smith, S. W., Daunic, A. P., & Taylor, G. G. (2007). Treatment fidelity in applied educational research: Expanding the adoption and application of measures to ensure evidence-based practice. *Education & Treatment of Children, 30*, 121–134.

Stein, M. K., & Wang, M. C. (1988). Teacher development and school improvement: The process of teacher change. *Teaching and Teacher Education, 8*, 5–35.

Stormont, M., & Reinke, W. M. (2012). Using coaching to support classroom-level adoption and use of interventions within school-wide positive behavioral interventions and supports. *Beyond Behavior, 21*, 11–19.

Sugai, G., & Horner, R. H. (2008). What we know and need to know about preventing problem behavior in schools. *Exceptionality, 16*, 67–77.

Sutherland, K. S., Wehby, J. H., & Copeland, S. R. (2000). Effect of varying rates of behavior specific praise on the on-task behavior of students with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders, 8*, 2–8.

Walker, B., Cheney, D., Stage, S., & Blum, C. (2005).Schoolwide screening and positive behavior support: Identifying and supporting students at risk of school failure. *Journal of Positive Behavior Interventions, 7*, 194–204.

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*, 398–410.

Ward, E., King, M., Lloyd, M., Bower, P., Sibbald, B., Farrelly, S., & Addington-Hall, J. (2000). Randomised controlled trial of non-directive counseling, cognitive-behaviour therapy, and usual general practitioner care for patients with depression. *British Medical Journal, 321*, 1383–1388.

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.

Wills, C. E., & Holmes-Rovner, M. (2006). Integrating decision making and mental health interventions research: Research directions. *Clinical Psychology: Science and Practice, 13*, 9–25.