



Published in final edited form as:

Prev Sci. 2013 December ; 14(6): 593–605. doi:10.1007/s11121-012-0343-9.

Understanding Real-World Implementation Quality and “Active Ingredients” of PBIS

Lauren E. Molloy, Julia E. Moore, Jessica Trail, John James Van Epps, and Suellen Hopfer
The Pennsylvania State University

Abstract

Programs delivered in the “real world” often look substantially different from what was originally intended by program developers. Depending on which components of a program are being trimmed or altered, such modifications may seriously undermine the effectiveness of a program. In the present study, these issues are explored within a widely used school-based, non-curricular intervention, Positive Behavioral Intervention and Supports. The present study takes advantage of a uniquely large dataset to gain a better understanding of the “real-world” implementation quality of PBIS, and to take a first step toward identifying the components of PBIS that “matter most” for student outcomes. Data from 27,689 students and 166 public primary and secondary schools across seven states included school and student demographics, indices of PBIS implementation quality, and reports of problem behaviors for any student who received an office discipline referral (ODR) during the 2007-2008 school year. Results of the present study identify three key components of PBIS that many schools are failing to implement properly, three program components that were most related to lower rates of problem behavior (i.e., three “active ingredients” of PBIS), and several school characteristics that help to account for differences across schools in the quality of PBIS implementation. Overall, findings highlight the importance of assessing implementation quality in “real-world” settings, and the need to continue improving understanding of how and why programs work. Findings are discussed in terms of their implications for policy.

Keywords

implementation quality; positive behavior supports; office discipline referrals

Understanding Real-World Implementation Quality and “Active Ingredients” of PBIS As schools continue to face problems of bullying, substance use, and other deviant behaviors, teachers and administrators are increasingly realizing the need for prevention. To reduce problem behaviors, effective programs must be identified, disseminated, and well implemented. In contrast to research trials, “real world” program implementers often make modifications due to limited time and resources; thus, the program delivered is often quite different from what was intended (Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005). Such modifications are problematic because they may “cut back” on components critical to a

program's effectiveness. Thus, a growing literature has emerged on *implementation quality*, or the degree to which program delivery adheres to the original program model. By examining implementation quality of a program's components in relation to outcomes, a better understanding can be achieved of which components "matter most". Yet research relating implementation quality to program outcomes is primarily conducted within controlled efficacy and effectiveness trials. In short, research is needed to examine how implementation quality relates to program outcomes in the real world across a diverse range of programs and settings.

Implementation Quality

Implementation quality matters because programs delivered with high quality are more likely to produce the desired effects (Durlak & DuPre, 2008). Barriers to implementation quality (e.g., insufficient facilitator training or lack of administrative support) have been found to reduce program effect sizes to half or a third of what they would otherwise be (e.g., Domitrovich & Greenberg, 2000; Lipsey, 2009). In contrast to controlled trials, "real-world" program implementation often does not closely follow the original program and involves adaptations that have not been empirically validated. School-based programs are particularly unlikely to resemble the original program when delivered in the real world, with lower implementation quality than that seen in family prevention or treatment, or community-based programs (Dariotis, Bumbarger, Duncan, & Greenberg, 2008). In one study of seven school-based sites, none fully adhered to the program model (Dariotis et al., 2008), and Gottfredson and Gottfredson (2002) found overall low implementation quality across a national sample of schools delivering delinquency prevention programs. Clearly, more attention must be paid to real-world implementation of school-based programs.

Often, implementation quality is collapsed into one score to indicate how well a program was implemented overall (e.g., Durlak & Dupre, 2008; Elias, Zins, Graczyk, & Weissburg, 2003), masking the strengths and weaknesses of a given setting. If adherence to a program is high on some components but poor on others, an overall score makes the measure far less informative (Gottfredson & Gottfredson, 2002). Instead, measures of implementation quality should assess the *multiple components* that make up a program (Durlak & Dupre, 2008). This data is valuable as feedback for continuous quality improvement (Domitrovich & Greenberg, 2000), but is also critical as a first step toward unpacking the "black box" of how a program leads to outcomes (Fixsen, et al., 2005; Greenberg, 2010). By examining implementation quality of program components in relation to outcomes, effects can be more clearly explained (i.e., what went on to cause particular effects in a given setting) and the impact of particular components can be assessed (Elias et al., 2003). Often referred to as the program's *active ingredients*, the program components most related to outcomes are useful in unveiling the mechanisms of change underlying a program's effects. Identifying a program's active ingredients allows for informed decisions about the program components that should receive the bulk of available time and resources and that require strict adherence in order to achieve the intended effects (Michie, Fixsen, Grimshaw, & Eccles, 2009). Such efforts can be seen in several recent studies (e.g., Cane, O'Connor, & Michie, 2012; Michie, van Stralen, & West, 2011); the current study bolsters this work by examining active ingredients within a widely disseminated school-based program.

Positive Behavioral Intervention and Supports

Most implementation quality research has focused on interventions taught in classroom settings, similar to a reading or math curriculum (e.g., Life Skills Training; Botvin, Eng, & Williams, 1980). Far less attention has been paid to *non-curricular* programs, in which school-wide efforts are made to change the school context or climate (e.g. Olweus Bullying Prevention Program; Olweus, 1991). The current study examines the implementation quality of one such non-curricular intervention: Positive Behavioral Intervention and Supports (PBIS). Currently, nearly 18,000 primary and secondary schools (approximately 18% of all schools in the nation) across 49 states implement PBIS, representing an 80% increase from 2008 (Spaulding et al., 2008). PBIS uses a “systems approach”: it is a school-wide prevention strategy that establishes a positive school climate and the behavioral supports needed to reduce behavior problems and enhance academic performance (see Bradshaw, Koth, Thornton, & Leaf, 2009). At the universal level (the focus of this study), PBIS modifies school environments by establishing clear school rules and through improved systems (e.g., rewards, discipline) and procedures (e.g., continual use of data for decision-making) that promote positive change in staff and student behaviors.

PBIS is appealing to schools for several reasons. First, rather than a “one size fits all” package, PBIS is based around meeting each school’s unique needs, emphasizing continuous use of student outcome data to make school policy decisions, and establishing sustainable system supports (TA Center on PBIS, 2012). Second, PBIS takes a pro-active approach to defining and teaching a continuum of positive behavior support for all students. This is simpler than piecing together multiple student-specific behavioral management plans, and is backed by evidence that punishing misbehavior is less effective than teaching and reinforcing positive behaviors (TA Center on PBIS, 2012). Third, PBIS is evidence-based: randomized controlled trials have demonstrated positive effects of PBIS on perceptions of school safety (Horner et al., 2009), academic performance, school suspensions, and frequency of ODRs (Bradshaw, Mitchell, & Leaf, 2009). Given its documented effectiveness, theoretical grounding, and widespread use, the importance of examining PBIS implementation quality in real world settings is clear.

Student Outcomes: Office Discipline Referrals

The recommended and most widely used source of data for continual monitoring of PBIS effectiveness has been *Office Discipline Referrals* (ODRs; Sugai, Sprague, Horner, & Walker, 2000). Referral forms are filed each time a student is referred to the office for violating a school rule. ODRs are a practical data source: they are widely available and follow a standard format across schools. They are also empirically relevant: they have been linked to poor student outcomes such as school failure and juvenile delinquency (see Clonan, McDougal, Clark, & Davison, 2007). Substantial evidence suggests that different problem behaviors have distinct correlates, and a given intervention may be more effective for some behaviors than others (e.g., CSPV, n.d.). Thus, to identify the *active ingredients* of PBIS, implementation quality is examined in relation to three categories of ODRs representing problem behaviors commonly targeted by prevention efforts: aggression or violence, substance use or possession, and defiance.

Factors Related to Implementation Quality

Once *active ingredients* have been identified, it is important to understand factors that account for between-setting differences in their implementation quality so that strategies may be identified to address them. Factors at multiple levels (e.g., intervention, individual, and school characteristics) have been linked to implementation quality in school settings (Domitrovich et al., 2008). For instance, better organizational capacity and support (Gottfredson & Gottfredson, 2002), administrative and principal support (Dariotis et al., 2008), and a better working alliance among teachers and implementers (Wehby, Maggin, Moore, & Robertson, 2012) have been linked to higher implementation quality. However, much of this work remains theoretical; more research is needed to empirically identify school-level predictors of implementation quality.

The Present Study

The present study uses data from real-world delivery of PBIS across a diverse range of schools to address three aims. The first aim is to descriptively report real-world implementation quality for each of the seven program components. The second aim is to take a first step toward identifying the *active ingredients* of PBIS with respect to three problem behaviors (aggression, drug use, and defiance). The third aim is to test whether school-level factors help to account for between-school variation in implementation quality of the *active ingredients*. As these analyses are largely exploratory, a priori hypotheses are not proposed.

Methods

Participants and Procedures

Data were drawn from a longitudinal dataset (2005-2008) that includes the School-wide Information System (SWIS) and the Positive Behavior Supports (PBS Surveys) data collection systems from the Education and Community Supports (ECS) research unit at the University of Oregon. The sample comprises all schools that were actively using the SWIS during the three years of the study and that completed the School-wide Evaluation Tool (SET). The PBS dataset provides indices of school demographics and PBIS implementation quality (i.e., responses to the SET), and the SWIS dataset provides information about every ODR filed in a given school year.

Because student data is logged by ODRs, it is impossible to know if a student is “missing” in a given year because he left the school or because he did not commit any offenses. Also, participant identification numbers were not consistent across buildings in a school district. Thus, a substantial portion of the data’s “nestedness” (i.e., wave nested within student) cannot be accounted for (Raudenbush and Bryk, 2002). For these reasons, longitudinal examination of the student-level data would be both methodologically and conceptually problematic. Given this study’s focus on how problem behaviors relate to implementation quality, the third (i.e., last available) year of data (2007-2008) was selected for the present study, ensuring that every school had been implementing PBIS for at least two years. Thus far only one study has examined length of time to full implementation of PBIS, finding that

it took two to five years for the majority of schools to be implementing with fidelity (Harms, 2010). Thus, the third year of data should provide the most representative indices of implementation quality for each school, making it most appropriate for this study. In the remainder of the paper, only data from the third year will be reported. No data was available on students with no ODRs; thus, they are not included in student-level analyses or sample characteristics described below.

Sample—Data come from 166 primary (i.e., K-6th grade, 65%) and secondary public schools within 65 school districts across 7 states. Demographic indicators obtained from the National Center on Education Statistics (NCES; 2006-2007) indicate that 25.9% of schools were located in cities, 52.4% in suburbs, 5.4% in towns, and 13.3% in rural areas. Locale data was not available for the remaining 3%. Compared to present-day national statistics, suburban schools were somewhat over-represented in this sample (with the national percentage at 34.4%), while towns and rural areas were under-represented (with national percentages at 11.9% and 13.3%, respectively), and cities were about accurately represented (with the national percentage at 29%; NCES, 2010-2011). The mean number of students per participating school was $M = 543$ ($SD = 317$); 48% of students per school ($SD=28\%$) were eligible for free or reduced-price lunch (“FRL”); and 65% of schools were eligible for Title I support. These values are representative of national statistics (548 students is the average school size, 48% of students are eligible for FRL, and 67% of schools nation-wide are Title I eligible). The sample size of students used for analysis in the present study (i.e., students with an ODR in the 2007-2008 school year) was $N=27,689$ students (67% male). Among those with available ethnicity data (71% of the sample), 40% were White, 33% Black, 24% Latino, and 3% were Asian, Native American, or Pacific Islander. Students with an Individualized Education Plan (“IEP”) made up 6% of the sample.

Measures

Implementation quality—The *Schoolwide Evaluation Tool* (SET) is a 28-item multi-method, multi-informant site-based evaluation tool designed to assess the implementation quality of school-wide PBIS (Sugai, Lewis-Palmer, Todd, & Horner, 2001). Trained staff at each school administer and score the SET, which includes observations, interviews with administrators, students, and school staff, and products provided by the school (e.g., referral forms). The 28 SET items are scored on a 0-2 scale (0=not implemented, 1=partially, 2=fully). Consistent with standard procedure (Horner et al, 2004), the 28 items were aggregated into seven theoretical components of PBIS implementation quality. Construct validity of the SET has been established by correlations with a measure of school-wide behavior support implementation completed by faculty and staff; and internal validity has been demonstrated with high inter-correlations among factors of the SET ($r = .44$ to $.81$; Horner et al., 2004).

Past research has demonstrated test-retest reliability and internal consistency of the SET, with alphas above the standard cut-off for acceptable reliability ($\alpha > 0.7$) on most of the subscales (Horner et al., 2004). In the present data, several of the alpha values fell below this cutoff; possible reasons are noted in the limitations section. To maintain external validity, the subscales employed in standard practice were used here: *Expectations Defined* (defining

school-wide behavioral expectations; $M = 1.81$, $SD = .32$; 2 items, $r = .24$), *Expectations Taught* (teaching behavioral expectations to all students; $M = 1.78$, $SD = .31$; 5 items, $\alpha = .64$); *Reward System* (acknowledging students for meeting behavioral expectations; $M = 1.85$, $SD = .35$; 3 items, $\alpha = .80$); *Violation System* (correcting misbehaviors using a continuum of behavioral consequences; $M = 1.69$, $SD = .34$; 4 items, $\alpha = .39$); *Monitoring and Decision-Making* (gathering and using student behavioral data to guide decision-making; $M = 1.90$, $SD = .21$; 4 items, $\alpha = .55$); and *Management* (administrative leadership of school-wide practices; $M = 1.77$, $SD = .27$; 8 items, $\alpha = .46$). A correlation of $r = -.02$ between the two items making up the seventh program component, *District-Level Support*, provided no justification for aggregating the items. Thus, the two items were considered separately: *District-Level Support: Budget* (allocation of school budget money to maintaining school-wide behavior supports; $M = 1.81$, $SD = .59$) and *District-Level Support: Liaison* (administrator can identify an out-of-school liaison in the district or state; $M = 1.86$, $SD = .52$). Summary scores are created for each subscale as the percentage of possible points earned, and an overall implementation score is computed as the mean of the subscales (Horner et al., 2004). Mean overall SET score across the 166 schools was $M = .90$ ($SD = .11$, $\alpha = .83$). Within the present sample, 83% of schools met the criteria for implementing PBIS with fidelity as defined by the PBIS literature (full implementation on at least 80% of the SET items and on at least 80% of the items assessing *Expectations Taught*).

Office discipline referrals—Each time a student was referred to the office for a problem behavior, information about that office discipline referral (ODR) was recorded (e.g., date, time, location, and behavior). To ensure valid and reliable ODR data, schools needed to meet a series of requirements (e.g., attendance at a training workshop, adoption of a standardized ODR form) before using SWIS, the web-based information system for collecting ODR information. All participating schools agreed to adopt a mutually exclusive set of operationally defined problem behaviors for use on their referral forms (Clonan, McDougal, Clark, & Davison, 2007). The widespread use of ODRs makes them a practical and valuable standardized instrument across schools. Although there is some potential for teacher bias in documentation (e.g., due to variability in teachers' tolerance for misbehavior), the use of ODRs has received empirical support as a valid and reliable tool to monitor student problem behaviors for purposes of decision-making in schools (e.g., Irvin et al., 2004). ODRs have been empirically linked to poor student outcomes such as poor academic achievement, substance use, and teacher reports of classroom disruptiveness (e.g., Pas, Bradshaw, & Mitchell, 2011).

The present study focuses on three broad categories of problem behaviors; an aggregate count was computed per student of the number of ODRs received in each category. *Aggression and violence* refer to behaviors intended to threaten, intimidate, or harm others, such as abusive language, bullying, and possession of a weapon. Mean number of *aggression/ violence* ODRs (per student with at least one ODR of any type) was $M = 1.79$ ($SD = 3.30$), ranging from 0 to 100; of students on record with an ODR, over 60% had received at least one *aggression/ violence* ODR. *Defiance* refers to behaviors that violate school rules or disrupt classroom functioning, such as truancy, insubordination, or cheating. Defiance was the most common, with over 75% of offending students receiving at least one

defiance ODR (M = 4.00; SD = 7.62; ranging from 0 to 156). *Drug use or possession* ODRs were far less common: less than 3% of referred students had a *drug*-related ODR (M = .03; SD = .22; ranging from 0 to 5). *Defiance* ODRs moderately correlated with *aggression* ODRs ($r = .47, p < .001$), while *drug*-related ODRs only weakly correlated with *aggression* ($r = .03, p < .001$) and *defiance* ($r = .11, p < .001$).

The sample reported above represents approximately 32% of all students within the larger population of students from the 166 schools sampled. In other words, of all students attending the sampled schools (N = 86,887), 32% had received at least one ODR of any type in the 2007-2008 school year. Percentages per school ranged from 7% to 93%, with a mean of 32%. Considered by category, 19% of all students had at least one aggression ODR, 24% had at least one defiance ODR, and just under 1% had at least one drug-related ODR.

Analytic Plan

To get a descriptive picture of real-world PBIS implementation (aim 1), percentages of schools fully implementing each of the PBIS components are reported. To identify the *active ingredients* of PBIS (aim 2), multilevel models (students nested within schools) were computed, predicting frequency of each ODR type (aggression, defiance, and drug-related) for all students with at least one ODR (students without any ODRs were excluded from analyses). Predictors in these models included the school-level SET factors representing the PBIS core components, controlling for student- and school-level demographics. Student gender and IEP status were tested as moderators of the school-level variables. Last, linear regressions tested whether school characteristics (school size, school level, and percent FRL) predicted implementation quality of the active ingredients (aim 3). Analyses for each aim are discussed in further detail below.

Results

“Real World” Implementation Quality of PBIS

The first aim of the study was to explore the percentage of schools that achieved *full implementation* on each of the seven PBIS components. For each component, schools were categorized as achieving *full implementation* if they received a “2” (high implementation quality) on each of the items used to assess that component. As shown in Figure 1, the percentage of schools achieving full implementation varied by program component. Over two thirds of schools fully *Defined Expectations*, set up a *Reward System*, used disciplinary data for *Monitoring and Decision-Making* about school-wide efforts, and received *District-Level Support* as originally intended and found effective. Yet only 49% of schools properly *Taught Expectations*, 37% had a full quality *Violation System* in place, and 37% reported full quality *Management*.

Identifying the “Active Ingredients” of PBIS

The next step was to identify the “active ingredients” of PBIS: three problem behavior categories were examined in relation to the implementation quality of the seven program components. Multilevel models (in SAS Proc Mixed; Raudenbush & Bryk, 2002) tested whether the number of times a student was reported for aggressive, defiant, or drug-related

behaviors varied as a function of concurrent implementation quality of the seven PBIS components after controlling for student- (gender, IEP status) and school-level (school size, percent FRL, and school level) demographic variables. Interactions of student demographics with school-level variables (demographic and implementation quality) were also tested. To aid interpretation, all non-dichotomous variables in the hierarchical models were grand-mean centered.

Backward elimination was used to determine the final models. There is no “hard and fast rule” for determining a cutoff significance value for trimming terms when using backward elimination, and a look across published social science studies reveals cutoff significance values ranging from .05 to .20. Due to the large number of terms being tested and the consequent difficulty of detecting significant effects, we chose the more conservative significance value cutoff of .20 to reduce the risk of a type II error (i.e., incorrectly trimming a term that is truly related to the outcome). To maintain consistency across models, main effects of the demographic variables were never dropped. See Table 1 for parameter estimates from the final models.

Across all models, problem behaviors were more common within *secondary schools* than within primary schools. Higher *percentage of FRL students* minimized the difference in defiance between boys and girls, but was unrelated to aggressive or drug-related behaviors. Students were less likely to be reported for defiance or aggression in *larger schools*. Lastly, problem behaviors were most common among *boys* (aggression) and *youth with IEPs* (defiance). Interactions among control variables are reported in Table 1 but are not elaborated here.

Regarding implementation quality, results indicate that the “active ingredients” vary by outcome under consideration, and may differ by gender and IEP status. *Aggressive and violent ODRs* were less common in schools with high quality implementation on the *Reward System* ($\beta=-2.74$, $p<.01$) and *Violation System* ($\beta=-4.17$, $p<.001$), and these relations were most prominent among non-IEP youth ($\beta=2.46$, $p<.001$; $\beta=3.75$, $p<.001$, respectively; see Figure 2 for graphs of these interactions). Results were mixed for *Expectations Defined* and *District Support: Liaison*, with interactions suggesting effects in differing directions by IEP status and gender (see Table 1). Lastly, implementation quality of *Expectations Taught* and *Management* related to aggression in the unintended direction for both non-IEP and IEP youth ($\beta=-1.81$, $p<.01$; $\beta=-3.87$, $p<.001$, respectively). For *defiance*, students received fewer ODRs in schools with high implementation quality on *Expectations Taught* ($\beta=-9.24$, $p<.001$), *Reward System* ($\beta=-6.61$, $p<.01$), and *Violation System* ($\beta=-7.77$, $p<.001$). Once again, high implementation more strongly related to the defiant behaviors of non-IEP students ($\beta=8.28$, $p<.001$ for interaction of IEP with *Expectations Taught*; $\beta=5.15$, $p<.001$ for *Reward System*; and $\beta=5.47$, $p<.001$ for *Violation System*; see Figure 2). However, there were also components that related to defiance in the unintended direction: *Expectations Defined* ($\beta=12.47$, $p<.001$), *Monitoring and Decision-Making* ($\beta=9.67$, $p<.01$), *Management* ($\beta=9.68$, $p<.01$), and *District Support: Budget* ($\beta=1.75$, $p<.01$). Interaction effects indicate stronger relations to defiant behaviors for non-IEP students and boys (see Table 1). Lastly, results of the *drug use or possession* model indicated that *Expectations Taught* and *Reward System* were significantly related to ODRs in the desired direction. There was a significant

main effect of *Reward System* on drug-related ODRs ($\beta=-.07$, $p<.05$), and significant interactions of *Expectations Taught* with gender ($\beta=-.05$, $p<.01$) and IEP status ($\beta=.07$, $p<.05$), indicating stronger relations among males and non-IEP youth (see Figure 2). Drug-related ODRs were associated with *District Support: Budget* in the intended direction among non-IEP youth, but in the unintended direction among IEP youth ($\beta=-.03$, $p<.05$), and were associated with *Management* in the unintended direction for both genders ($\beta=-.05$, $p<.05$).

In sum, findings across models suggest that ODRs were lowest when *Expectations Taught*, *Reward System*, and *Violation System* were well implemented; relations between ODRs and other components of PBIS were mixed. Lastly, a relatively consistent pattern emerged suggesting weaker relations between implementation quality and ODRs among IEP youth.

School Characteristics Associated with Implementation Quality

To address the third aim, linear regression was used to test whether characteristics of schools predict implementation quality of the three active ingredients identified above. Parameter estimates are presented in Table 2, and Figure 3 presents mean implementation quality among schools characterized by high and low levels on the three demographic variables (i.e., one SD above or below the mean). *School level* (primary versus secondary school) significantly predicted implementation quality of *Expectations Taught* ($\beta=.18$, $p<.05$), and marginally significantly predicted quality of schools' *Reward Systems* ($\beta=.14$, $p<.10$): in both cases, implementation quality was higher in primary schools. *School size* was negatively predictive of *Reward Systems* ($\beta=-.33$, $p<.001$), and marginally negatively predictive of *Expectations Taught* ($\beta=-.16$, $p<.10$): smaller schools achieved higher implementation quality. Finally, *percent of students with FRL* negatively predicted *Expectations Taught* ($\beta=-.16$, $p<.05$), suggesting that higher SES schools achieve better implementation quality than low SES schools.

Discussion

Programs delivered in the “real world” often look substantially different from what was originally intended by program developers, which may undermine their effectiveness. As a result, empirical interest in implementation quality has recently spiked; unfortunately, it remains largely under-examined in real-world settings. To help schools make the best use of limited resources, important goals for the field include investigating which program components schools struggle with most, how implementation quality relates to program outcomes, and what factors contribute to implementation quality in the real world. Concurrent relations observed in this study are a valuable first step toward achieving these goals: results suggest three components of PBIS that the majority of schools are failing to implement properly, three components that significantly relate to student outcomes, and several school characteristics that help to account for between-school differences in PBIS implementation quality.

“Real World” Implementation Quality of PBIS

First, percentages of schools were identified that achieved “full implementation” on each of the seven theoretical PBIS program components. A school was considered to be “fully

implementing” a given component if it met SET standards for full implementation (not necessarily 100% adherence) on each item measuring that component. For instance, one indicator of a “fully implemented” *Reward System* requires “50% or more of students asked to indicate that they have received a reward for expected behaviors over the past two months”. On three of the components, schools achieving full implementation were in the minority: only 49% of schools properly *Taught Expectations*, 37% of schools had a full quality *Violation System*, and 37% of schools showed proper *Management*. In light of these findings, it is worth considering what can be done to help schools improve. For instance, are there alternative strategies that might achieve the same end goal but have a greater chance of successful implementation?

Identifying the “Active Ingredients” of PBIS

Next, implementation quality of the seven program components was examined in relation to three categories of ODRs, as a first step toward identifying *active ingredients* of real-world PBIS. Results are suggestive of three active ingredients. High quality implementation of the *Reward System* predicted lower rates of all three behavioral categories; high implementation of the *Violation System* predicted fewer defiance and aggression ODRs; and properly *Teaching Expectations* predicted fewer defiance and drug-related ODRs. It is intuitive that these three components would relate to student behaviors. Students were more likely to meet expectations and less likely to violate those expectations when they knew what those expectations were and of the associated rewards and sanctions. While other components take place at the administrative and planning levels, these three components provide clear guidelines and concrete contingencies for student behaviors. These findings represent concurrent relations; certainly, longitudinal studies will be needed to reach more definitive conclusions. Yet the consistency of these relations across the three models provides compelling evidence that schools benefit from placing a priority on teaching behavioral expectations and setting up reward and violation systems linked to those expectations. These findings are a useful first step toward filling in the “black box” of PBIS and its relation to outcomes.

Several PBIS components related to problem behaviors in the unexpected direction. One such component for which implementation quality was most consistently related to problem behaviors in the unexpected direction was *Management*. It is unlikely that findings should be misinterpreted to indicate *negative effects* of PBIS. Given that the relations are concurrent, one possibility is that the relation of ODRs to *Management* and other higher-level planning components reflect a tendency for schools to improve the implementation quality of those components *in response to* problematic rates of ODRs. For instance, one of the items used to evaluate implementation quality of *Management* asks: “Does the school improvement plan list improving behavior support systems as one of the top 3 school improvement plan goals?” In other words, *improving* positive behavior supports likely becomes a high priority when defiance, aggression, and drug use become a prevalent cause for concern. In general, *Management* (in this context) is about *setting up* positive behavior supports: high *Management* scores may represent a school that is still in a “set-up” phase of PBIS implementation.

Lastly, several of the relations between PBIS components and problem behaviors were moderated by IEP status: relations were generally *weakest* among youth with an IEP. This could be because such school-wide practices are redundant with youths' IEPs: for instance, IEPs may already involve concrete contingencies for meeting and violating expectations. However, it may also be the case that these types of school-wide, non-curricular programs are failing to reach special-needs students. Based on the moderating effects of IEP status observed consistently in the present study, it seems more research would be useful to understand: a) why PBIS implementation quality is less related to IEP-youths' behaviors, and b) whether alterations should be made to PBIS to increase its reach to these youth.

School Characteristics Associated with Implementation Quality

Findings suggest that *primary schools*, *smaller schools*, and *higher-SES schools* achieved higher quality implementation, consistent with past research (e.g., Brown, Feinberg, & Greenberg, 2010; Domitrovich et al., 2008). Though the mechanisms remain largely unknown, all three of these factors were predictive of the implementation quality of *Expectations Taught*, and school level and school size predicted implementation quality of the *Reward System*. One obvious explanation is that higher SES schools simply have more resources to achieve higher implementation quality, and will be in a better position to prioritize and successfully execute intervention programs (Domitrovich et al., 2008). Similarly, with a smaller student to staff ratio, staff at smaller schools may have more time and capacity to closely monitor individual student behaviors and respond appropriately. Lastly, drug use does not typically emerge until adolescence, some degree of increased "deviancy" is relatively normative during adolescence, and defiant and aggressive behaviors in secondary school are often more advanced than similarly classified behaviors in primary school. The higher prevalence of problem behaviors among secondary school students may also translate to more thinly spread resources and greater challenges to implementation. Whether reflecting insufficient resources or a poor fit between PBIS implementation processes and the needs or abilities of low-SES, larger, or secondary schools, the lower implementation quality in these schools will need further investigation.

Implications for Policy

Given the widespread use of PBIS, awareness of the most common strengths and weaknesses of its implementation is critical. In the present study, two of the three "active ingredients" of PBIS were also found to be two of the most poorly implemented components. Providing schools with the resources and knowledge to monitor their own performance may be one way to address this issue; this is supported by evidence that continual monitoring has been successful at improving intervention implementation and outcomes (e.g., Bickman, 2008; Fagan, Arthur, Hanson, Briney, & Hawkins, 2011) and is increasingly being recommended (e.g., Becker & Domitrovich, 2011). In developing future programs, researchers must carefully consider how to increase the likelihood of successful implementation in real-world settings. Given evidence of lower implementation quality in lower-SES, secondary, and larger schools, future research and policy efforts should work toward providing and studying the impact of alternate implementation processes that may be a better fit for these settings.

It is also critical that the path from programs to outcomes be well understood, especially given the reality of limited resources. Outside of randomized controlled trials, examining associations between implementation quality and student behaviors is the only way to learn which program components work best. In turn, the field can begin to gain insight into a program's "black box", providing valuable information for future program improvement and development. Identifying a program's *active ingredients* within real-world settings allows researchers and practitioners to make informed decisions about how best to maximize cost-efficiency. For instance, school administrators can ensure that resources are being spent on the most effective program components (e.g., those most likely to improve administrator, teacher, and student behaviors). Finally, a greater understanding is needed of factors that pose challenges to implementation quality and how to counteract those challenges (e.g., Tibbits, Bumbarger, Kyler, & Perkins, 2010). For instance, re-structuring large schools into smaller "pods" may allow faculty and staff to be responsible for a more reasonable number of students, which has been found to foster more positive teacher-student relationships, more positive school climates, and fewer conduct problems among students (e.g., Barber & Olsen, 2004; Bethea, 2012).

Strengths, Limitations, & Future Directions

The data examined in the present study provided a unique and valuable opportunity to address this study's research aims. PBIS is one of a small set of widely disseminated programs, now implemented in 18% of US schools. This study is among the first to use implementation science to further understanding of what works in real-world settings. The large sample allows for examination of variability across schools and comprehensive assessment of implementation quality. Given the standardization of data collection procedures and diverse sample of schools, these data are a useful representation of the broader population of schools implementing PBIS.

In the development of future databases like the SWIS, several additions would address limitations of the present dataset. First, data collection procedures should track year of implementation. It is currently unknown how long PBIS to reach "full implementation" once introduced; additionally, year of implementation provides an important context for interpreting program effects (or lack thereof) and allows investigation of the link between year of implementation and program effectiveness. Second, given limitations of participant identification procedures in the present dataset, only contemporaneous relations could be examined. For future studies to determine the direction of these relations, participant identification numbers should be consistent across school years and buildings. Lastly, no data were available for students who had not received an ODR in the 2007-2008 school year. Consequently, analyses predicting student-level ODRs tested relations with implementation quality among students *who had received at least one ODR* – in other words, those already offending – but did not allow for prediction of *non-offending*. PBIS may have helped to eliminate all instances of problem behavior for some students by the 2007-2008 school year; it is further possible that different components of PBIS would have significantly related to ODRs had non-offending students been included. When possible, data should allow distinction between students who left the school versus those who did not commit offenses; most ideal would be to collect data about all students in a school.

As noted, measures of internal reliability on the SET subscales fell below the standard cut-off ($\alpha > .70$) within the present study. One reason for low reliability with these data is ceiling effects – several of the program components displayed a limited amount of variability across schools. Further limiting reliability levels, several subscales included only two or three items. Low reliability does not affect the likelihood of a type I error (false positive), but it does increase the likelihood of a type II error (false negative). In other words, the significant effects reported in the present study needed to be robust in order to emerge as significant; however, with greater internal consistency of the subscales, additional effects may have emerged that could not be detected here. Future investigations should explore other explanations for low reliability, and should consider adding more items and/or a wider range of response options to the existing measures to allow for more precise detection of differences across schools.

Concluding remarks—To achieve changes in public health, effective programs must be identified, disseminated, and implemented with high quality. The present investigation is one of a limited set of studies that has examined implementation quality and its relation to outcomes within a real world setting. As evidence accumulates that relations between implementation quality and outcomes hold across multiple programs and various outcomes, the field can increasingly shift its focus toward understanding *which components* of a program affect *which outcomes*. To the extent that programs' *active ingredients* can be identified, the field will be better equipped to improve upon programs and make the best use of limited resources.

Acknowledgments

Data for this study were obtained from the OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports. This research was supported by the Office of Special Education Programs US Department of Education (H326S980003). Opinions expressed herein are those of the authors and do not necessarily reflect the position of the US Department of Education, and such endorsements should not be inferred.

References

- Barber BK, Olsen JA. Assessing the transitions to middle and high school. *Journal of Adolescent Research*. 2004; 19:3–30. doi: 10.1177/0743558403258113.
- Becker KD, Domitrovich CE. The conceptualization, integration, and support of evidence-based interventions in the schools. *School Psychology Review*. 2011; 40:582–589.
- Bethea, KR. A cross-case analysis of the implementation and impact of smaller learning communities in selected S.C. public middle schools. Retrieved from ProQuest Information & Learning. 2012. (AAI3454674)
- Bickman L. A measurement feedback system (MFS) is necessary to improve mental health outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2008; 47:1114–1119. doi: 10.1097/CHI.0b013e3181825af8. [PubMed: 20566188]
- Botvin GJ, Eng A, Williams CL. Preventing the onset of cigarette smoking through life skills training. *Preventive Medicine*. 1980; 9:135–143. doi: 10.1016/0091-7435(80)90064-x. [PubMed: 7360727]
- Bradshaw CP, Koth CW, Thornton LA, Leaf PJ. Altering school climate through school-wide Positive Behavioral Interventions and Supports: Findings from a group-randomized effectiveness trial. *Prevention Science*. 2009; 10:100–115. doi: 10.1007/s11121-008-0114-9. [PubMed: 19011963]
- Bradshaw CP, Mitchell MM, Leaf PJ. Examining the effects of schoolwide positive behavioral interventions and supports on student outcomes: Results from a randomized controlled effectiveness

- trial in elementary schools. *Journal of Positive Behavior Interventions*. 2010; 12:133–148. doi: 10.1177/1098300709334798.
- Brown LD, Feinberg ME, Greenberg MT. Determinants of community coalition ability to support evidence-based programs. *Prevention Science*. 2010; 11:287–297. doi: 10.1007/s11121-010-0173-6. [PubMed: 20352332]
- Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. *Implementation Science*. 2012; 7(37) doi: 10.1186/1748-5908-7-37.
- Center for the Study and Prevention of Violence. Blueprints for violence prevention. (n.d.)Retrieved from <http://www.colorado.edu/cspv/blueprints/index.html>
- Clonan SM, McDougal JL, Clark K, Davison S. Use of office discipline referrals in school-wide decision making: A practical example. *Psychology in the Schools*. 2007; 44:19–27. doi: 10.1002/pits.20202.
- Dariotis JK, Bumbarger BK, Duncan LG, Greenberg MT. How do implementation efforts relate to program adherence? Examining the role of organizational, implementer, and program factors. *Journal of Community Psychology*. 2008; 36:744–760. doi: 10.1002/jcop.20255.
- Domitrovich CE, Bradshaw CP, Poduska JM, Hoagwood K, Buckley JA, Olin S, Ialongo NS. Maximizing the implementation quality of evidence-based preventive interventions in schools: A conceptual framework. *Advances in School Mental Health Promotion*. 2008; 1(3):6–28. doi: 10.1080/1754730x.2008.9715730.
- Domitrovich CE, Greenberg MT. The study of implementation: Current findings from effective programs that prevent mental disorders in school-aged children. *Journal of Educational & Psychological Consultation*. 2000; 11:193–221. doi: 10.1207/s1532768xjepc1102_04.
- Durlak JA, DuPre EP. Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*. 2008; 41:327–350. doi: 10.1007/s10464-008-9165-0. [PubMed: 18322790]
- Elias MJ, Zins JE, Graczyk PA, Weissburg RP. Implementation, sustainability, and scaling up of social-emotional and academic innovations in public schools. *School Psychology Review*. 2003; 32:303–319.
- Fagan AA, Hanson K, Briney JS, Hawkins JD. Sustaining the utilization and high quality implementation of tested and effective prevention programs using the communities that care prevention system. *American Journal of Community Psychology*. 2012; 49:365–377. doi: 10.1007/s10464-011-9463-9. [PubMed: 21809149]
- Fixsen, DL.; Naoom, SF.; Blase, KA.; Friedman, RM.; Wallace, F. *Implementation research: A synthesis of the literature*. University of South Florida; Tampa, FL: 2005.
- Gottfredson DC, Gottfredson GD. Quality of school-based prevention programs: Results from a national survey. *Journal of Research in Crime and Delinquency*. 2002; 39:3–35. doi: 10.1177/002242780203900101.
- Greenberg MT. School-based prevention: Current status and future challenges. *Effective Education*. 2010; 2:27–52. doi: 10.1080/19415531003616862.
- Harms, ALS. A three-tier model of integrated behavior and learning supports: Linking system-wide implementation to student outcomes. Retrieved from ProQuest Information & Learning. 2011. (AAI3433110)
- Horner RH, Sugai G, Smolkowski K, Eber L, Nakasato J, Todd AW, Esperanza J. A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions*. 2009; 11:133–144. doi: 10.1177/1098300709332067.
- Horner RH, Todd AW, Lewis-Palmer T, Irvin LK, Sugai G, Boland JB. The School-Wide Evaluation Tool (SET): A research instrument for assessing school-wide positive behavior support. *Journal of Positive Behavior Interventions*. 2004; 6:3–12. doi: 10.1177/10983007040060010201.
- Irvin LK, Tobin TJ, Sprague JR, Sugai G, Vincent CG. Validity of office discipline referral measures as indices of school-wide behavioral status and effects of school-wide behavioral interventions.

- Journal of Positive Behavior Interventions. 2004; 6:131–147. doi: 10.1177/10983007040060030201.
- Lipsey MW. The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims & Offenders*. 2009; 4:124–147. doi: 10.1080/15564880802612573.
- Michie S, Fixsen D, Grimshaw J, Eccles M. Specifying and reporting complex behaviour change interventions: The need for a scientific method. *Implementation Science*. 2009; 4(40) doi: 10.1186/1748-5908-4-40.
- Olweus, D. Bully/victim problems among schoolchildren: Basic facts and effects of a school based intervention program. In: Rubin, KH.; Pepler, DJ., editors. *The development and treatment of childhood aggression*. Erlbaum; Hillsdale, NJ: 1991. p. 411-448.
- OSEP Positive Behavioral Interventions and Supports. What is school-wide PBIS?. 2012. Retrieved from <http://www.pbis.org/school/default.aspx>
- Pas ET, Bradshaw CP, Mitchell MM. Examining the validity of office discipline referrals as an indicator of student behavior problems. *Psychology in the Schools*. 2011; 48:541–555. doi: 10.1002/pits.20577.
- Raudenbush, SW.; Bryk, AS. *Hierarchical linear models: Applications and data analysis methods*. Vol. Vol. 1. Sage; Newbury Park, CA: 2002.
- Spaulding, S.; Horner, R.; May, S.; Vincent, C. Evaluation brief: Implementation of school-wide PBS across the United States. Nov. 2008 Retrieved from [http://www.pbis.org/evaluation/evaluation_briefs/nov_08_\(2\).aspx](http://www.pbis.org/evaluation/evaluation_briefs/nov_08_(2).aspx)
- Sugai, G.; Lewis-Palmer; Todd, AW.; Homer, RH. *School-wide Evaluation Tool (Version 2.0)*. Educational and Community Supports, University of Oregon; Eugene, OR: 2001.
- Sugai G, Sprague JR, Horner RH, Walker HM. Preventing school violence: The use of office discipline referrals to assess and monitor school-wide discipline interventions. *Journal of Emotional and Behavioral Disorders*. 2000; 8:94–101. doi: 10.1177/10634266000800205.
- Tibbitts MK, Bumbarger BK, Kyler SJ, Perkins DF. Sustaining evidence-based interventions under real-world conditions: Results from a large-scale diffusion project. *Prevention Science*. 2010; 11:252–262. doi: 10.1007/s11121-010-0170-9. [PubMed: 20229358]
- Wall, RM. Factors influencing development and implementation of local wellness policies. Retrieved from ProQuest Dissertations & Theses. 2011. (M.S. 1499500)

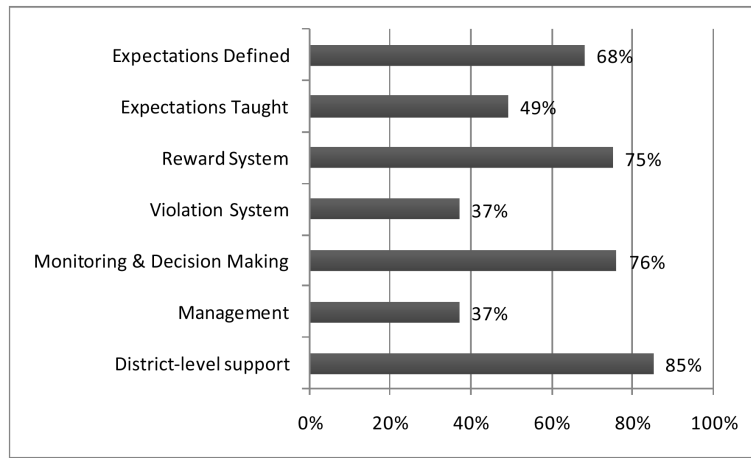


Figure 1. Percentage of schools fully implementing each of the PBIS program components.

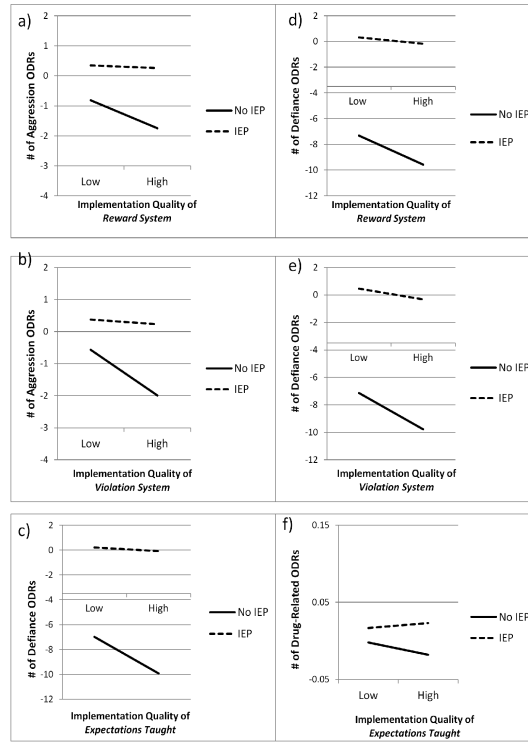


Figure 2. Interactions of IEP status with *Reward System* (panel a) and *Violation System* (panel b) predict number of *aggression* ODRs per student; with *Expectations Taught* (panel c), *Reward System* (panel d), and *Violation System* (panel e) predict number of *defiance* ODRs per student; with *Expectations Taught* (panel f) predicts number of *drug-related* ODRs per student.

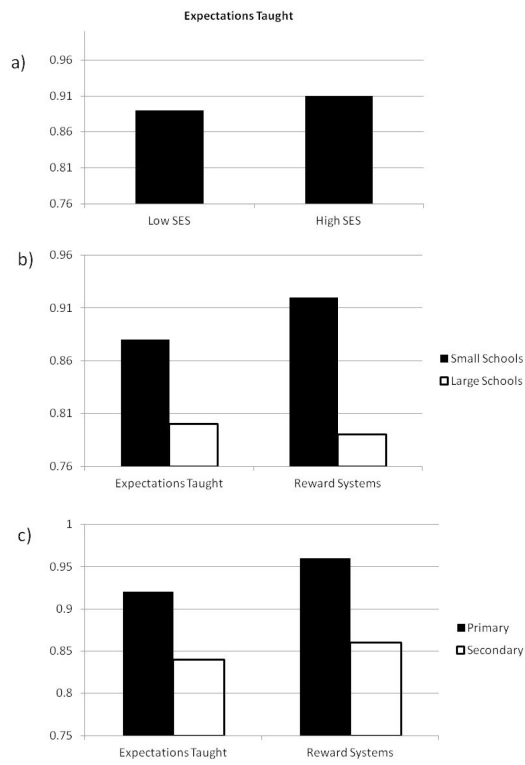


Figure 3. Implementation quality of PBIS “active ingredients” by SES (panel a), school size (panel b), and school level (panel c).

Table 1

Predicting Student ODRs from School and Student Demographics and Implementation Quality of PBIS Program Components

Fixed Effects	Aggression		Defiance		Substance Use	
	Coef.	SE	Coef.	SE	Coef.	SE
Intercept	-1.28	1.26	-8.45*	3.4	-0.01	0.05
SCHOOL-LEVEL						
<i>School Demographics</i>						
Pct free/reduced lunch	0.11	0.49	0.66	0.92	<.01	0.01
School Level (Primary/ Secondary)	1.39***	0.30	4.38***	0.67	.04***	0.01
School Size	<-.01***	<.01	<.01	<.01	<.01	<.01
<i>Implementation Quality of Program Components</i>						
Expectations Defined	2.21*	1.02	12.47***	2.19	0.05	0.04
Expectations Taught	2.08*	1.03	-9.24***	2.34	-0.05	0.04
Reward System	-2.74**	0.94	-6.61**	2.25	-.07*	0.03
Violation System	-4.17***	0.89	-7.77***	2.03		
Monitoring and Decision-Making			9.67**	3.48		
Management	5.39***	1.34	9.68**	3.08	0.07	0.04
Budget Support			1.75**	0.6	0.02	0.01
District/ State Liaison	0.32	0.30	-0.24	0.53		
STUDENT-LEVEL						
<i>Student Demographics</i>						
Gender	0.91**	0.29	0.55	1.02	0.11***	0.01
Gender * Percent FRL			-1.10**	0.37		
Gender * School Level	-0.40***	0.09				
Gender * School Size	<.01*	<.01	<.01***	<.01		
IEP Status	1.58	0.85	8.52***	2.3	0.03	0.04
IEP Status * Percent FRL	0.49	0.31				
IEP Status * School Level	-1.16***	0.18	-1.66***	0.42		
IEP Status * School Size	<.01***	<.01	<.01**	<.01		
<i>Implementation Quality by Student Demographics</i>						
Gender * Expectations Defined			1.31*	0.53		
Gender * Expectations Taught					-0.05**	0.02
Gender * Reward System	0.39	0.25	-0.94	0.52	-0.02	0.01
Gender * Violation System						
Gender * Monitoring			1.76	1.03		
Gender * Management	0.56	0.31			-0.05*	0.02

Fixed Effects	Aggression		Defiance		Substance Use	
	Coef.	SE	Coef.	SE	Coef.	SE
Gender * Budget Support						
Gender * District/ State Liaison	-0.27**	0.10	-0.36	0.23		
IEP Status * Expectations Defined	-2.53***	0.66	-10.48***	1.31	-0.06	0.03
IEP Status * Expectations Taught	-1.81**	0.62	8.28***	1.49	0.07*	0.03
IEP Status * Reward System	2.46***	0.60	5.15***	1.48		
IEP Status * Violation System	3.75***	0.59	5.47***	1.43		
IEP Status * Monitoring			-9.02***	2.43		
IEP Status * Management	-3.87***	0.86	-7.15***	2.11		
IEP Status * Budget Support			-1.27**	0.41	-0.03*	0.01
IEP Status * District/ State Liaison	-0.47*	0.19				
Random Effects						
Between School						
Intercept	1.57		7.54		<.01	
Residual Error	8.36		47.10		.04	

* Note. $p < .05$;

** $p < .01$;

*** $p < .001$

Table 2

Predicting the Implementation Quality of the Three “Active Ingredients” of PBIS from School Characteristics

Independent Variables	Implementation Quality of “Active Ingredients”		
	Expectations Taught	Reward System	Violation System
School Level (Primary/ Secondary)	.18*	.14 ⁺	-.12
School Size	-.16 ⁺	-.33***	.12
Percent Free/ Reduced-Price Lunch	-.16*	-.04	-.02
R ²	.289	.398	.169

* Note: p<.05,

⁺ p<.10