The Relationship Between School-Wide Implementation of Positive Behavior Intervention and Supports and Student...

Article *in* Journal of Positive Behavior Interventions · June 2015

DOI: 10.1177/1098300715590398

CITATIONS
5

READS
196

4 authors:

Karen E Childs
University of South Florida
18 PUBLICATIONS **684** CITATIONS
SEE PROFILE

Donald Kincaid
University of South Florida
18 PUBLICATIONS **701** CITATIONS
SEE PROFILE

Heather Peshak George
University of South Florida
20 PUBLICATIONS **74** CITATIONS
SEE PROFILE

Nicholas Gage
University of Florida
47 PUBLICATIONS **332** CITATIONS
SEE PROFILE

Hammill Institute on Disabilities

Journal of Positive Behavior Interventions
1–11

DOI: 10.1177/1098300715590398
jpbi.sagepub.com
SAGE

The Relationship Between School-Wide Implementation of Positive Behavior Intervention and Supports and Student Discipline Outcomes

Karen Elfner Childs, MA[1], Don Kincaid, EdD[1], Heather Peshak George, PhD[1], and Nicholas A. Gage, PhD[2]

Abstract

School-Wide Positive Behavior Interventions and Supports (SWPBIS) is a systems approach to supporting the social and emotional needs of all children utilized by more than 21,000 schools across the nation. Data from numerous studies and state projects' evaluation reports point to the impact of SWPBIS on student outcomes (office discipline referrals [ODRs], in-school suspensions [ISSs], out-of-school suspensions [OSSs]) and the possible relationship between implementation fidelity and those student outcomes. With data from 1,122 Florida schools, this study used a longitudinal design to examine the associations between the total score and 10 subscale scores on the Benchmarks of Quality (BoQ), a validated SWPBIS implementation fidelity measure, and school-level behavioral outcomes: ODRs, ISSs, and OSSs. Results of these analyses found a decreasing trend across all three behavioral outcomes, and schools having higher BoQ total scores realized lower ODRs and had corresponding fewer ISSs and OSSs. Of the 10 subscales, the Classroom was negatively and significantly associated with ODRs and OSSs, whereas the BoQ Data Entry Plan was positively and significantly associated with ODRs at initial status and across time after controlling for school-level characteristics (e.g., size, number of years of implementation). The implications of the findings for SWPBIS assessment and intervention in the classroom are discussed.

Keywords

school-wide, intervention(s), data analysis, studies, positive behavior, support(s)

School-Wide Positive Behavior Interventions and Supports (SWPBIS) is a systems approach to establishing the social culture and behavioral supports needed for all children in a school to achieve both social and academic success. Over the past 30 years, a number of studies have documented the effectiveness of SWPBIS. This growing body of research supports improvements in disciplinary behavior, school climate, organizational health, student bullying behavior and peer victimization, and academic achievement (Bradshaw, Koth, Thornton, & Leaf, 2009; Bradshaw, Mitchell, & Leaf, 2010; Horner, Sugai, & Anderson, 2010; Sadler & Sugai, 2009; Simonsen et al., 2012; Waasdorp, Bradshaw, & Leaf, 2012). As a result, there have been more than 21,000 schools across the country trained in SWPBIS (Horner, 2013). Three states have more than 60% of schools involved in SWPBIS implementation; 9 states have more than 40%, and 16 states have more than 30% of their schools implementing SWPBIS (Sugai & Simonsen, 2012). These numbers represent the expansive efforts of state and district leadership teams in scaling up SWPBIS implementation to continue to realize positive student outcomes (Bradshaw, Leaf, & Debnam, 2007; McIntosh et al., 2013) and the need to build capacity to sustain those efforts at the local level.

SWPBIS is a systems framework for schools to establish social and behavior supports to increase academic gains and reduce problem behavior across all students using evidence-based practices (Sugai & Horner, 2006). The primary features of SWPBIS include (a) capitalizing on the prevention of problem behavior, (b) teaching appropriate social behavior and skills, (c) acknowledging appropriate behavior, (d) using a multitiered approach to instruction/intervention that matches behavior support intensity to student need, (e) using data-based problem solving, and (f) investing in

[1]University of South Florida, Tampa, USA
[2]University of Florida, Gainesville, USA

Corresponding Author:
Karen Elfner Childs, University of South Florida, 13301 Bruce B. Downs Boulevard, Tampa, FL 33612, USA.
Email: kchilds2@usf.edu

Action Editor: Kathleen Lane

systems that support evidence-based practices (George, Kincaid, & Pollard-Sage, 2009; www.pbis.org).

Schools that are effective in SWPBIS implementation (i.e., implementing the critical elements) are often referred to as high implementers or as implementing with fidelity (George & Childs, 2012). Fidelity of implementation is "the extent to which the delivery of an intervention adheres to the protocol or program model originally developed" (Mowbray, Holter, Teague, & Bybee, 2003, p. 315). The extent to which schools implement SWPBIS with fidelity is of value to researchers and practitioners alike. Researchers examine the evidence for SWPBIS effectiveness and use fidelity measures to distinguish between schools that implement and those that do not (Tobin et al., 2012). Practitioners at the school level focus on creating a safe and orderly school environment by examining SWPBIS implementation fidelity measures as a means to realign resources and design interventions to promote positive student outcomes.

Many validated instruments exist that are designed to assess the level of implementation fidelity specifically for Tier 1 behavior support and provide valuable information to assist a team in developing an action plan for improved program design. In fact, most schools trained in SWPBIS are using some type of instrument to assess the level of implementation fidelity on their school campus (Tobin et al., 2012). Two of the most widely used instruments are the (a) School-Wide Evaluation Tool (SET; Horner et al., 2004) and the (b) Benchmarks of Quality (BoQ; Kincaid, Childs, & George, 2010). The SET focuses on the initial implementation activities completed by a trained outside evaluator. Recent research on the SET suggests it to be very reliable at the elementary school level and may be most appropriate for schools initiating SWPBIS (Pas & Bradshaw, 2012; Vincent, Spaulding, & Tobin, 2010). The BoQ is a self-report measure with widespread use both nationally and internationally that has proven to be a reliable, valid, efficient, and useful instrument for measuring the fidelity of implementation at the Tier 1 level of SWPBIS in individual schools (Cohen, Kincaid, & Childs, 2007; George & Childs, 2012).

Both measures are proven tools for use in SWPBIS evaluation and have their advantages and disadvantages. The advantage of the SET is that it is an external evaluation tool, which is less susceptible to the bias of a self-assessment tool like the BoQ. However, external evaluation can be costly and less efficient for implementation scale-up (e.g., personnel time, training, travel). The information provided by the SET is oriented toward reporting on outcomes (existence of posters, knowledge of expectations), rather than feedback about the implementation process which can lead to specific action planning afforded by the BoQ. The BoQ is organized around 10 critical domains (i.e., subscales) reflecting the essential components of SWPBIS: (a) SWPBIS team, (b) faculty commitment, (c) effective procedures for dealing with discipline, (d) data entry and analysis plan established, (e) expectations and rules developed, (f) reward/recognition program established, (g) lesson plans for teaching expectations/rules, (h) implementation plan, (i) classroom systems, and (j) evaluation. The original validation study of the BoQ (Cohen et al., 2007) provided evidence of strong internal consistency (overall $\alpha = .96$), interrater reliability ($r = .87$), and test–retest reliability ($r = .94$). An assessment of concurrent validity through comparison with the SET (Horner et al., 2004) found moderate correlations (.51). A supplementary assessment of concurrent validity was also conducted with data from 720 schools completing both the SET and the BoQ with results showing a significant relationship: $r = .52$ and $p < .0001$. A factor analysis conducted by Childs, Kincaid, and George (2011) led to a revision of the measure that included removal of seven weak items including the entire "crisis" section and the addition of seven previously piloted classroom items that had demonstrated to form one consistent factor. The inclusion of items related to the actions of individual teachers within their classrooms was identified as one of the most prevalent suggestions for improvements during the first 5 years of the BoQ's increasingly widespread utilization.

Fidelity of implementation is an essential variable in assessing the impact of SWPBIS on student performance (George & Childs, 2012). In their examination of the evidence base for SWPBIS, Horner et al. (2010) found growing evidence of the association between implementation of Tier 1 SWPBIS and a reduction in reports of problem behavior as well as improved perception of school safety. In a randomized controlled effectiveness trial with 37 elementary schools, Bradshaw et al. (2009) demonstrated that schools were able to implement SWPBIS with fidelity, and office discipline referrals (ODRs) and the proportion of students receiving out-of-school suspensions (OSSs) declined. Horner et al. (2009) found the improvements in the perception of school safety within a randomized control trial with 63 elementary schools implementing SWPBIS. In related research, Pas and Bradshaw (2012) found a significant relationship between implementation of SWPBIS and student academic and truancy outcomes in a study involving 474 schools from 24 districts in Maryland.

More recently, Pas, Waasdorp, and Bradshaw (2014) examined how the level of SWPBIS implementation related to student outcomes and school-level contextual factors. Results indicated that higher implementation fidelity was associated with higher math and reading achievement as well as lower truancy. The researchers identified school contextual factors related to implementation levels and outcomes. The subscale scores of the SET representing the key features were examined, with no significant outcomes found. The total score for the BoQ was analyzed, with no significant relationship identified with outcomes. Subscale scores for the BoQ were not analyzed in this study. Although this study may provide some initial

support for the relationship between implementation and outcomes, it stopped short of examining whether the specific elements on the BoQ are related to student outcomes. This warrants an analysis of whether there are, in fact, particular critical features that are necessary or sufficient to produce changes in student outcomes (OSS, in-school suspension [ISS], and ODR).

Simonsen et al. (2012) attempted to address the relationship between fidelity of implementation and student outcomes by analyzing historical school-level data from Illinois. The authors modeled longitudinal differences in the number of ODR, ISS, and OSS by fidelity status, defined as a SET score of 80 or above. Results suggested that schools implementing with fidelity also had corresponding lower rates of student outcomes, but the difference in ODRs was not statistically significant. In addition, the models only examined overall fidelity of implementation as a dichotomous yes/no and did not examine whether or not SET subscales predicted decreases in student outcomes.

Examination of the relationship between critical features and student outcomes is further complicated by the literature related to stages of implementation (Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005). It may be that particular features are more or less critical for outcomes at different stages of SWPBIS implementation (e.g., exploration, installation, full implementation, innovation, etc.). For example, the stage of implementation may fluctuate across years of implementation. Although a compilation of features may be important such as represented in a total score, the individual features may not be equally weighted and indicate where a team should invest support at different points in the implementation process (Coffey & Horner, 2012). From a practical standpoint, the examination of the individual features, critical elements, or various components of SWPBIS implementation as identified by self-reported fidelity measures and their relationship to student outcomes may be of value for developing more effective and efficient implementation support.

To assess the relationship between both overall fidelity and the various components of SWPBIS implementation and negative student outcomes, we examined the longitudinal relationship between the BoQ and school-level reports of discipline outcomes, including ODRs, ISS, and OSS. Specific research questions guiding the study were as follows:

Research Question 1: Is there a decrease in the frequency of student discipline outcomes across time for schools implementing SWPBIS?

Research Question 2: Is the BoQ total score related to differences in school-level discipline outcomes at initial status (intercept) and across time (slope) after controlling for school-level characteristics (e.g., school size, number of years implementing SWPBIS)?

Research Question 3: Are the BoQ subscale scores related to differences in school-level discipline outcomes at initial status (intercept) and across time (slope) after controlling for school-level characteristics (e.g., school size, number of years implementing SWPBIS)?

Method

Study Design, Participants, and Setting

This study used a longitudinal design to examine the associations between the BoQ total score and subscale scores and school-level behavioral outcomes. The investigation included 4 years of data from 1,122 elementary, middle, and high schools collected between the 2010–2011 and 2013–2014 school years. All schools were collaborating with Florida's Positive Behavior Support: A Multi-Tiered System of Supports (FLPBS:MTSS) Project and submitted each of the three necessary evaluation elements: School Profile (demographic information), School-Wide Benchmarks of Quality (implementation fidelity measure), and the Outcome Data Summary (student discipline data). The participating schools included 724 elementary schools, 248 middle schools, and 150 high schools. School enrollment ranged from 81 to 7,756 students ($M = 844$; $SD = 461.45$); the percentage of students on free and reduced-price lunch ranged from 1% to 100% ($M = 54.88\%$; $SD = 31.79\%$). The average number of years implementing SWPBIS was 3.71 years ($SD = 0.98$). For use as a predictor in the growth model analysis, each school's total years of experience were summed across years (2 years, 3 years, 4 years, and 5 years of experience summed to 14 years), resulting in an average of 11.37 years ($SD = 8.71$).

Participating school teams completed 3 days of training that included lecture, team activities, and videotapes of Florida schools to assist in visualizing implementation on their campus and completed a comprehensive action plan to guide implementation activities. This Tier 1/Universal Level SWPBIS training is standard for all schools participating with FLPBS:MTSS Project and addresses the following topics: teaming, developing expectations and rules, system for teaching appropriate behavior, reward system, effective discipline procedures (definitions, effective responses to problem behaviors, forms and data collection, coherent office referral process), data analysis, staff commitment, planning, and evaluation. The Tier 1 training also includes an application of SWPBIS strategies for the classroom containing identification and teaching of classroom rules, routines, and procedures as well as practices for responding to appropriate and inappropriate behavior. Ongoing technical assistance and coaching vary depending upon the level of support provided by the district ranging from monthly coaches' meetings with on-site support and annual "booster" trainings to districts that provide only

Table 1. Average Benchmarks of Quality Scores for Schools.

	All (N = 1,122)		Elementary (n = 724)		Middle (n = 248)		High (n = 150)	
Scale	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
BoQ Total	77.82	15.17	80.52	13.47	74.57	15.45	69.77	18.59
PBS Team	86.83	15.21	88.29	13.35	85.96	14.14	83.26	17.78
Faculty Commitment	70.66	21.82	71.92	20.13	68.41	21.31	66.50	23.48
Effective Procedures	87.94	13.39	89.02	12.32	87.00	13.18	87.22	13.37
Data Entry Plan	71.76	19.86	73.07	18.44	72.12	18.27	70.17	19.15
Expectations	87.57	15.16	90.96	11.36	84.99	15.11	80.18	20.30
Reward Program	75.52	19.37	79.38	16.35	73.10	17.72	65.17	23.72
Lesson Plans	69.92	25.15	74.84	21.27	64.55	25.88	57.25	28.32
Implementation Plan	67.99	22.79	72.28	19.70	65.61	21.48	56.66	25.33
Evaluation	72.73	19.86	76.32	17.82	69.54	18.94	64.84	21.55
Classroom	82.32	17.67	86.81	14.12	77.06	17.86	73.36	20.54

Note. BoQ = Benchmarks of Quality; PBS = Positive Behavior Support.

minimal support to coaches and schools with no additional professional development opportunities.

Measures

BoQ. The BoQ (Cohen et al., 2007; Kincaid et al., 2010) is a psychometrically strong evaluation instrument broadly used to assess implementation fidelity at the Tier 1/universal level of SWPBIS using a 53-item rating scale. BoQ items are organized and scored around 10 critical elements along with a total score. The BoQ was the selected measure for this study as it is part of the standard protocol for FLPBS:MTSS Project comprehensive evaluation and was readily available.

Table 1 portrays the average fidelity of implementation scores for schools across time included in this study. The average BoQ total score was 78% (*Mdn* = 80%), with elementary schools implementing more critical features than high schools. The subscale with the highest percentage of implementation was Effective Procedures, while the subscale with the lowest percentage implementation was Implementation Plan. Overall, only 13% of the schools never achieved the 70% fidelity threshold over the 4 years, while 28% varied in their implementation, meaning 1 year their BoQ total score was below 70%, but above 70% the following year. For information about how the 70% threshold was established, see the initial BoQ validation study (Cohen et al., 2007).

School Profile. Schools provide demographic data (e.g., student enrollment, percent of students on free and reduced lunch) on an annual basis using the School Profile. School profile data used in this study included school type (e.g., elementary school), school enrollment size, and Title I status. In addition, we included the number of years each school had been implementing SWPBIS, defined as the number of years working directly with the FLPBS:MTSS Project. For analyses, we recoded school type into a dichotomous variable of elementary and secondary schools.

Outcome Data Summary. This summary collected at the close of the school year includes various outcome data including number of ODRs, days of ISS, and days of OSS. Florida schools follow guidelines established by the Florida Department of Education (FLDOE) for identifying and documenting ISS and OSS. FLDOE also defines and requires reporting of serious behavior incidences (e.g., felony offenses). Individual school districts and schools make decisions about defining other incidences resulting in ODRs and suspensions. Discipline data for this study included the number of ODRs, OSS, and ISS for each school.

Data Collection Procedures

After a school team participates in initial SWPBIS training, coaches/facilitators receive training to accurately complete and submit all evaluation requirements of FLPBS:MTSS Project, including completion of the BoQ and the Outcome Data Summary. This professional development occurs during district-coordinated Coaching 101 training facilitated live or web-based or via periodic district-level coaches' meetings. Additional resources to support the completion of evaluation instruments are available on the Project's website (http://flpbs.fmhi.usf.edu), including videos describing how to complete the instruments, PowerPoint files, and evaluation directions provided for each school throughout the year.

All data are submitted to FLPBS:MTSS Project via the web-based SWPBIS evaluation system by the coach/facilitator for each participating school utilizing a unique school identifier and encrypted password. The School Profile data are entered in October during the fall midyear evaluation

period, and the BoQ and Outcome Data Summary data are entered during May and June during the end-year evaluation period. The BoQ is scored in about an hour by the school team's coach/facilitator with input from team members using the Scoring Guide which contains operational definitions of the possible scores for every item. Each item has a maximum value between 1 and 4 points for a maximum total of 107 points (see George & Childs, 2012).

The coach/facilitator for each school completes the School Profile and Outcome Data Summary forms that include identifying the school enrollment, number of discipline referrals, and suspensions; coaches frequently receive assistance from other school personnel (i.e., school administrator and/or data clerk) to obtain accurate data for these forms. The source of enrollment and discipline data for each school is unique with most Florida school districts utilizing distinctively different systems for recording and managing student data.

All data are entered into Florida's PBIS Evaluation System and maintained in a single database to increase consistency. To safeguard accuracy, the database is equipped with error checking to identify duplications, outliers, and missing entries, with local school personnel verifying and/or correcting data identified as outliers. Only schools with each data element represented were included in the analysis.

Analytic Approach

To address the research questions, we calculated growth models for each behavioral outcome and modeled the relationship between the behavioral outcomes and the school average BoQ total score and BoQ subscale scores. We conducted growth models using structural equation modeling, where the intercept and slope are latent factors (Preacher, Wichman, MacCallum, & Briggs, 2008). We calculated models for the total and subscale scores separately because the total score is based on the sum of the subscales, and we assumed the multicollinearity would result in unstable model estimates (Raudenbush & Bryk, 2002). We modeled BoQ data as a time-invariant average across time (a) to assess the impact of overall fidelity of implementation on discipline outcomes (i.e., the 4-year average represents a school's general implementation), and (b) the BoQ scores moved up and down by only a few points for most schools, making interpretation difficult if modeled as time variant. In addition to BoQ total and subscale scores, we modeled time-variant and time-invariant covariates to account for school-level differences beyond fidelity of implementation scores. The growth models were as follows:

$$\text{Discipline}_{it} = \pi_{0i} + \pi_{1i}(\text{YEAR}_{it}) + \sum \pi_{it}\mathbf{T}_{it} + \sum \pi_i \mathbf{F}_i + \sum \pi_i \mathbf{S}_i + \zeta_i + \varepsilon_{it}$$

where Discipline is ODR, OSS, or ISS for school i at school year t, π_{0i} is the average discipline outcome at time 0, π_{1i} is a linear function of each school's discipline outcome by school year, **T** is a vector of time-varying covariates, including school size and Title I status, **F** is a vector of schools' average BoQ scores, **S** is a vector of time-invariant school characteristics, including school type and years of experiences implementing SWPBIS, and ζ_i and ε_i are error terms for school and time, respectively. The model assumed linear growth because implementation fidelity and years implementing should decrease discipline outcomes over the 4 years of data used in this study. Model fit statistics were calculated to assess how well the data fit our growth models. Model fit was assessed based on sufficient comparative fit index (CFI) and Tucker–Lewis index (TLI) values defined as > .95, and root mean squared error of approximation (RMSEA) ≤ .06 (Hu & Bentler, 1999).

Missing data were not present at the individual school and variable level, but 13.9% of schools did not submit data for all four school years. We can accurately calculate a growth model under these circumstances by using full information maximum likelihood estimation (Enders & Bandalos, 2001). All models were conducted in the LAVAAN package (Rosseel, 2012) in R (R Development Core Team, 2014).

Following What Works Clearinghouse (WWC) standards, we used the Benjamini–Hochberg (BH) false discovery rate correction (Benjamini & Hochberg, 1995) for multiple hypotheses testing to protect against Type I error across the models. The correction was included because the three behavioral outcomes are representative of the same domain (i.e., school-level behavior incidents). The BH correction was calculated in Microsoft Excel based on the formula given in Benjamini and Hochberg (1995). For ODRs, the minimal p value was .018 or less, $p < .021$ for ISS, and $p < .020$ for OSS.

Results

Descriptive Statistics

Table 2 shows means and standard deviations for each of the student discipline outcomes by school type. Descriptive results suggest that there was a decreasing trend across all three discipline outcomes. For example, ODR decreased by an average of approximately six ODRs per year. An exception to the decreasing trend was a slight increase in OSS and ISS from the 2012–2013 to the 2013–2014 school years. Large differences in the frequency of ODR, ISS, and OSS were present between elementary and secondary schools, which may have been related to school size as the average size of elementary schools was 670 students, while the average school size was 1,277 for secondary schools.

Table 2. School-Level Discipline Outcomes per 100 Students for Schools Implementing SWPBIS Across Time.

	2010–2011		2011–2012		2012–2013		2013–2014	
Discipline outcome	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
All schools								
ODR	107.80	70.16	101.03	70.73	86.61	69.66	84.35	70.76
ISS	37.28	37.59	34.89	34.93	28.74	34.85	29.65	38.30
OSS	41.90	41.27	38.69	38.70	36.24	36.56	37.41	41.30
Elementary school								
ODR	33.51	31.65	32.42	33.50	30.25	31.66	29.61	30.06
ISS	4.61	8.29	4.29	9.55	4.12	8.87	3.43	6.55
OSS	10.66	13.68	10.45	13.86	10.84	13.37	10.84	16.31
Middle school								
ODR	148.59	98.66	146.42	99.51	123.33	96.06	125.21	111.41
ISS	62.20	57.97	58.15	53.93	47.84	51.92	50.87	71.37
OSS	69.06	70.22	66.69	69.87	62.09	62.02	66.21	71.52
High school								
ODR	141.31	80.18	124.24	79.17	106.24	81.25	98.24	70.81
ISS	45.02	46.52	42.22	41.32	34.25	43.77	34.66	36.99
OSS	45.97	39.92	38.94	32.36	35.80	34.30	35.17	36.07

Note. SWPBIS = School-Wide Positive Behavior Interventions and Supports; ODR = office discipline referral; ISS = in-school suspension; OSS = out-of-school suspension.

Growth Models

The first step in growth modeling was the estimation of the unconditional model to (a) evaluate the intercept and slope of the dependent variables absent of all predictors, (b) confirm our linearity assumption, and (c) calculate an intraclass correlation coefficient (ICC). The unconditional models for all three dependent variables are presented in Table 3. The dependent variables were not transformed and can be directly interpreted. The average number of ODR at the intercept was 71, but the negative slope value suggests that schools decreased the number of ODR by approximately 4 each year. The results were very similar for ISS and OSS, with the average number of incidents 28 and 22, respectively. Both suspension variables decreased over time, with a larger annual decrease for ISS. The significant covariance for ODR and ISS indicates that as time increases, both discipline outcomes decrease. To assess the proportion of variance explained by time for the dependent variables, we calculated an ICC for each using their respective variance estimates (σ^2 and τ_{00}). The ICC was 0.03 for ODR, 0.01 for OSS, and 0.03 for ISS, suggesting time accounted for less than 5% of the total variance for all three discipline variables.

Next, we calculated full predictor models for each of the discipline outcomes, and the results are presented in Table 4. Model 1 for each of the discipline outcomes describes the relationship between the BoQ Total score and the outcomes, controlling for all school-level covariates. Model 1 results for ODR indicate that the BoQ Total score has a statistically significant, negative effect on the intercept, suggesting that, at the first measurement period, schools with higher BoQ scores had lower ODRs. However, BoQ Total did not predict any differences in slope change. The covariance statistics for ODR indicate that there was a significant negative relationship between time and each discipline outcome, suggesting that schools with higher BoQ scores start out with lower ODR and that the gap at the first measurement period remains consistent across time because the slope values are equivalent across schools across time. Covariate results for the ODR model indicate that elementary schools have, on average, 108 less ODR referrals than secondary schools at the intercept, while schools with more years of experience have more ODRs. Model 1 results for the intercept parameter for ISS and OSS were similar to those found for ODRs. Schools with higher BoQ scores had correspondingly fewer ISS and OSS. Results for the slope value were different for ISS, where results suggest that as BoQ scores increased each year, so too did the frequency of ISS. However, the increase was less than 0.1 ISS per 100 students for each unit increase on the BoQ, indicating a very small increase across time.

Model 2 for each of the discipline outcomes examined the predictive relationship of each of the BoQ subscale scores. For both ODR and OSS, the BoQ Classroom subscale predicted a statistically significant decrease in the frequency of each discipline outcome, meaning the more schools implemented effective SWPBIS practices in classrooms, the more likely they were to have fewer ODR and OSS. The only other subscale that significantly predicted differences at initial status was the BoQ Data Entry Plan,

Table 3. Unconditional Growth Models for School-Level Discipline Outcomes per 100 Students.

	ODR		OSS		ISS	
Fixed effects	γ	*SE*	γ	*SE*	γ	*SE*
Intercept (π_{0i})	71.09*	2.45	28.17*	1.39	21.67*	1.26
Slope (π_{1i})	−3.93*	0.069	−.082*	0.37	−1.35*	0.40
Covariance						
Intercept X						
Slope	−421.16*	62.41	−22.27	20.30	−124.57*	21.79
Variance						
Intercept (σ^2)	5,696.90	279.14	1,614.26	96.01	1,392.70	77.30
Slope (τ_{00})	192.61	23.30	8.67	8.42	50.32	9.08
Model fit						
χ^2	2,917.65[a]		2,157.39[a]		2,654.02[a]	
CFI	0.98		0.99		0.97	
TLI	0.98		0.99		0.97	
RMSEA	0.04		0.03		0.04	
LogLik	−18008.70		−16057.49		−16176.96	

Note. ODR = office discipline referral; OSS = out-of-school suspension; ISS = in-school suspension; *SE* = standard error; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; LogLik = log likelihood.
[a]Six degrees of freedom.
*$p < .05$.

which had a significant positive relationship. This suggests that schools implementing data entry with fidelity also reported more ODRs. No subscales predicted any differences across time (slope).

Discussion

This analysis is an example of how to examine student outcome data relative to the level of implementation at a statewide level while controlling for school-level differences. First, an examination of descriptive data indicated that there is a decreasing trend in the frequency of student discipline (ODR, ISS, and OSS) outcomes across time for schools implementing SWPBIS. Yet time only accounted for 5% of the variance for all three measures, suggesting little change across time. Second, growth modeling indicated that, related to implementation fidelity, ODR, ISS, and OSS outcomes were similar across time, indicating that fidelity did not predict differences in growth trajectories. However, schools with higher BoQ scores start out with lower ODR, ISS, and OSS rates compared with schools that have lower BoQ scores. At the end of 4 years, the same difference remains between schools: higher implementing schools were still better in comparison with lower performing schools. Finally, the results indicate that only the BoQ subscale scores for the Classroom subscale (negative correlation) and the Data Entry Plan (positive correlation) were related to differences in school-level discipline outcomes at initial status (intercept) and across time (slope) after controlling for school-level characteristics (e.g., school size, number of years implementing SWPBIS).

The results for Research Questions 1 and 2 are puzzling. It is important to know that SWPBIS produces a decreasing trend in discipline responses across years for implementing schools but is perplexing to see that while there is a difference between discipline incidents each year with high implementing schools always being lower on average than low implementing schools, the slope of change is the same for both across 4 years. From these data, it appears that implementation fidelity is critical for a sustained decrease in discipline incidents but that higher fidelity does not produce faster change as measured by the slope or rate of change. In other words, there appears to be an immediate drop in discipline incidents when SWPBIS is implemented with fidelity and that the drop is sustained across time but does not decrease at a rate any different than schools implementing with less fidelity. These results are worthy of further examination as they may point to the need to consider the impact of implementation on outcomes. Perhaps implementation fidelity most reliably produces decreases in discipline outcomes during early stages of implementation but is not sufficient to produce changes in the slope or rate of change for targeted schools. These results are interesting and worthy of greater study with assistance from our colleagues well versed in implementation science.

The finding that subscale scores for the Data Entry Plan (positive correlation) and the Classroom (negative correlation) predict differences in school-level discipline outcomes at initial status (intercept) and across time (slope) is not unexpected. The positive relationship between high scores on the Data Entry Plan subscale and higher ODR, ISS, and OSS reports is most likely due to the increased efficiency,

Table 4. Growth Models Predicting the Relationship Between the Benchmarks of Quality and School-Level Discipline Outcomes per 100 Students.

	ODR				ISS				OSS			
	Model 1		Model 2[a]		Model 1		Model 2[a]		Model 1		Model 2[a]	
Fixed effects	γ	*SE*	γ	*SE*	γ	*SE*	γ	*SE*	γ	*SE*	γ	*SE*
Intercept (π_{0i})	178.00	11.30	176.25	18.07	75.49	6.27	76.87	10.07	90.99	6.83	87.31	10.98
Elementary	−108.03*	4.47	−100.79	4.60	−52.26*	2.48	−50.27*	2.58	−51.84*	2.72	−48.31*	2.79
BoQ Total	−0.58*	0.13			−0.27*	0.07			−0.43*	0.08		
Data Entry Plan			0.49	0.17			0.11	0.09			0.21	0.10
Classroom			−0.67*	0.22			−0.22	0.12			−0.63*	0.13
Years Experience	0.57*	0.22	0.44	0.22	0.32*	0.12	0.27	0.12	0.44*	0.13	0.35*	0.13
Slope (π_{1i})	4.03	4.50	−0.53	7.10	−7.12	2.61	−8.85	4.14	2.71	2.45	6.57	3.86
Elementary	5.81*	1.76	5.67*	1.81	2.12	1.03	1.94	1.06	0.94	0.96	0.78	0.99
BoQ Total	−0.09	0.05			0.08*	0.03			0.00	0.03		
Years Experience	0.08	0.08	0.08	0.08	−0.02*	0.05	−0.02	0.05	−0.03	0.04	−0.03	0.04
Time-variant covariates												
Size 10	0.00	0.00	0.00	0.00	−0.01*	0.00	−0.01*	0.00	−0.01	0.00	−0.01	0.00
Title 10	6.82*	2.94	6.65*	2.94	5.84*	1.75	5.85*	1.76	7.18*	2.07	7.09*	2.06
Size 11	−0.01*	0.00	−0.01*	0.00	−0.01*	0.00	−0.01*	0.00	−0.01*	0.00	−0.01*	0.00
Title 11	8.33*	2.47	7.86*	2.47	4.02*	1.47	3.88*	1.48	6.78*	1.61	6.40*	1.61
Size 12	−0.02*	0.00	−0.02*	0.00	−0.01*	0.00	−0.01*	0.00	−0.01*	0.00	−0.01*	0.00
Title 12	5.41	2.61	4.91	2.62	4.22*	1.39	4.00*	1.40	4.66*	1.41	4.15*	1.41
Size 13	−0.02*	0.00	−0.02*	0.00	−0.01*	0.00	−0.01*	0.00	−0.02*	0.00	−0.02*	0.00
Title 13	3.22	3.39	2.45	3.41	3.70	2.01	3.29	2.04	4.54*	1.73	3.90	1.75
Covariance												
Intercept X												
Slope	−263.42*	49.18	−260.38*	48.27	−78.40*	17.86	−77.31*	17.71	−0.92	17.82	3.22	17.47
Pseudo R^2												
Intercept	0.51		0.54		0.47		0.49		0.45		0.50	
Slope	0.04		0.07		0.06		0.08		0.02		0.11	
Model fit												
χ^2	3,778.30[b]		3,855.90[c]		2,837.30[b]		2,873.80[c]		2,873.80[b]		3,418.10[c]	
CFI	0.96		0.96		0.97		0.97		0.97		0.94	
TLI	0.94		0.93		0.96		0.96		0.96		0.91	
RMSEA	0.06		0.05		0.05		0.05		0.03		0.05	
LogLik	−56677.70		−96344.40		−54792.80		−94479.60		−94479.60		−94609.50	

Note. A table with all results for all BoQ subscales can be obtained from the authors. ODR = office discipline referral; ISS = in-school suspension; OSS = out-of-school suspension; *SE* = standard error; BoQ = Benchmarks of Quality; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; LogLik = log likelihood.

[a]Only subscales that significantly predicted discipline outcomes were included; all other subscales were not statistically significant. [b]Fifty degrees of freedom. [c]Eighty-six degrees of freedom.

*ODR *p* value < .018; ISS *p* value < .021, and OSS *p* value < .020; significance levels based on Benjamini–Hochberg.

and potentially accuracy, of data entry. It would not be surprising to see lower rates of discipline incidents from schools that also did not have an effective and efficient plan for data entry. Higher scores on Data Entry Plan most likely reflect a greater commitment and capacity for accurate reporting of incidents.

One explanation for the significance of the Classroom subscale is that it may be possible for a school to receive a high score on many items on the BoQ yet not see corresponding changes in student outcomes. Many of the items on the BoQ address the foundational activities and initial products of the SWPBIS process. These distal activities and products may only be remotely related to the measurement of student outcomes. In fact, many activities (discipline process is developed, problem behaviors defined, data system developed, etc.) occur prior to or at the beginning of the SWPBIS process, although the measurement of student outcomes may not be completed until the end of the year. However, classroom indicators of SWPBIS are not likely to be scored high on the initial BoQ assessment. The classroom items measure whether classrooms are implementing the SWPBIS system. In other words, the SWPBIS process starts with foundational work, the process is developed and implemented across the school, and then the process extends to the classroom setting as the last area of impact. Perhaps the assessment of the degree of implementation of SWPBIS in the classroom is an assessment of the degree to which the SWPBIS system has really permeated all the necessary school settings.

An alternative interpretation is that student outcomes may not be significantly impacted until SWPBIS has been extended and implemented with fidelity in the classroom. This would not be surprising as students spend most of their day in the classroom setting, and the classroom teacher has

the most opportunity to teach, acknowledge, and respond to positive and problematic behaviors. However, Tier 1 supports at the classroom level (i.e., supports across all classrooms) often receive the least amount of attention and tend to present the greatest inconsistencies in SWPBIS implementation (Newcomer, 2009; Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008). In fact, in most schools, ODRs from the classroom (which generate the ISS and OSS numbers) account for more than 50% of all of the ODRs for a school. If the majority of staff are not consistent in implementing SWPBIS (including within individual classrooms), the school is bound to lack the ability to achieve a high level of implementation and desired outcomes. If schools are to maximize efforts at the Tier 1 level, these systems of support in the classroom must be addressed. Therefore, the effective implementation of SWPBIS within all classrooms should have an immediate and significant impact on student outcome data.

Limitations

We note the limitations of using statewide data for this analysis. First, the data utilized in this analysis were from one state (Florida). While schools and students in Florida are likely similar to all other states, the training and support provided by the state's SWPBIS system may differ considerably from other states. Future studies should examine the relationship between classroom implementation and student outcomes in other states that may implement SWPBIS with differing training and support processes.

Second, the collection of implementation and outcome data across an entire state and from more than a thousand schools provides many possible threats to the fidelity of the submitted data. All the data submitted are self-report data and are susceptible to inaccurate and unreliable reporting by schools. For instance, while the state determines the definitions of what constitutes an OSS or ISS, there is little control of the implementation of consistent procedures across all schools. Infractions that result in ODRs may also vary from school to school along with the fidelity of their use. There can be inconsistencies from school to school in the types of incidents that result in a discipline referral, which may result in variability of the data. The evaluators collecting these statewide data spend considerable time reviewing, correcting, and prompting district and school personnel to submit accurate and timely data. However, online data collection is subject to inadvertent key strokes and other threats to accuracy that require constant vigilance and correction by evaluators.

Another constraint with statewide level data is the need to balance the quantity of data schools report with its value for the purpose of evaluation. As such, we note a limitation of the OSS data utilized in our analysis. We examined the total days of suspension because those are the data provided by the school but acknowledge that the number of suspension events and the number of students contributing to the suspensions would provide a more comprehensive representation of the metric. Similarly, the models were constrained to only four time points, limiting the addition of nonlinear growth terms in the models. Future research should include more time points and assess potential deceleration of discipline data across time.

A final limitation within the data gathered for this analysis is a realization that most schools that submit implementation and outcome data to the state project do so to document their success with SWPBIS. The fact that the median BoQ score for all schools submitting data for this sample is 80% indicates that our analyses are likely skewed toward more successful schools and that we have limited data from less successful implementers. Perhaps if the data were more normally distributed, we would have identified greater or stronger associations between variables.

Future Directions

The results of this study indicate the critical role that the fidelity of school-wide application of SWPBIS, particularly at the classroom level, has on the achievement of valued student outcomes. Furthermore, research may be necessary to define and clarify the relationship between classroom implementation and outcomes. Is the implementation of SWPBIS in the classroom an *indicator* of the extent to which SWPBIS is infused within the entire school and thus lays the foundation for significant student outcomes? On the other hand, is the implementation of SWPBIS in the classroom a *prerequisite* for achieving significant student outcomes? The answer to these questions may have a significant impact on the application of SWPBIS. If implementation in the classroom is an *indicator*, then the field of SWPBIS may need to better measure the implementation of classroom SWPBIS in a more systematic and accurate manner. If implementation of SWPBIS in the classroom is a *prerequisite*, then training and technical assistance on the development and implementation of a school-wide system must expand explicitly to target classroom environments versus other school settings.

In addition to these critical questions, further analysis of large data sets may tell us whether there are particular classroom scores or full scale scores on the BoQ that are related to outcome changes. What score is necessary on the entire BoQ or on particular subscales before a school will likely see consistent and significant student outcome changes? Alternatively, are there threshold scores (80%, 85%, 90%, etc.) or relative increases (10% 25%, 50%, etc., increase in fidelity from last year) on the BoQ or on subscales that are related to student outcomes? The size of our data set did not allow for an analysis of the influence of the year of implementation on implementation levels and student outcomes.

In addition, our available data from schools only reflect the year of implementation and not the phases of implementation (Fixsen, Blase, Metz, & Van Dyke, 2013). It will be critical to assess whether there are clear impacts on student outcomes by the phase of a school's implementation of SWPBIS.

Finally, are student outcome changes or sustainability impacted by other factors? Kincaid, Childs, Wallace, and Blase (2007) identified some barriers and facilitators of SWPBIS implementation from both high- and low-implementing schools. Although a wide range of barriers and facilitators emerged from school teams, the most frequently cited critical areas included district and state support, staff buy-in, data systems, inconsistent implementation, a reward system, time, staff turnover, and differing philosophies in the school. McIntosh et al. (2013) identified factors related to sustainability of SWPBIS implementation including two school-level factors—school priority and use of data—and two district-level factors—district priority and capacity building. Future research may also want to investigate how some of these critical areas also interact with classroom implementation to produce student outcomes. Such studies should prove useful in further defining an effective and efficient training and technical assistance support model to scale up SWPBIS implementation from approximately 21,000 schools to more than 100,000 schools nationwide.

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 development of this article was supported in part by Grant #H326S130004 from the Office of Special Education Programs, U.S. Department of Education for OSEP Center on Positive Behavioral Interventions and Supports (www.pbis.org). Opinions expressed herein do not necessarily reflect the position of the U.S. Department of Education, and such endorsement should not be inferred.

References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*, 289–300.

Bradshaw, C. P., Koth, C. W., Thornton, L. A., & Leaf, P. J. (2009). Altering school climate through school-wide Positive Behavioral Interventions and Supports: Findings from a group-randomized effectiveness trial. *Prevention Science*, *10*, 100–115.

Bradshaw, C. P., Leaf, P., & Debnam, K. (2007, July). *Project target*. Paper presented at the Maryland State SWPBIS Conference, Baltimore, MD.

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). 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 Behavioral Interventions*, *12*, 161–179.

Childs, K. E., Kincaid, D., & George, H. P. (2011). *The revised school-wide PBIS Benchmarks of Quality (BoQ)*. Available from http://www.pbis.org

Coffey, J., & Horner, R. H. (2012). The sustainability of school-wide positive behavior support. *Exceptional Children*, *78*, 407–422.

Cohen, R., Kincaid, D., & Childs, K. (2007). Measuring school-wide positive behavior support implementation: Development and validation of the Benchmarks of Quality (BoQ). *Journal of Positive Behavior Interventions*, *9*, 203–213. doi:10.1177/10983007070090040301

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*, 430–457. doi:10.1207/S15328007SEM0803_5

Fixsen, D. L., Blase, K. A., Metz, A., & Van Dyke, M. (2013). Statewide implementation of evidence-based programs. *Exceptional Children*, *79*, 213–230.

Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa: Louis de la Parte Florida Mental Health Institute, University of South Florida.

George, H. P., & Childs, K. E. (2012). Evaluating implementation of schoolwide behavior support. *Preventing School Failure*, *56*, 197–206. doi:10.1080/1045988X.2011.645909

George, H. P., Kincaid, D., & Pollard-Sage, J. (2009). Primary tier interventions and supports. In W. Sailor, G. Dunlap, G. Sugai & R. Horner (Eds.), *Handbook of positive behavior support* (pp. 375–394). Lawrence, KS: Issues in Clinical Child Psychology.

Horner, R. H. (2013, June 12). *Implementing SWPBIS with quality, equity and efficiency*. Available from www.pbis.org

Horner, R. H., Sugai, G., & Anderson, C. M. (2010). Examining the evidence base for school-wide positive behavior support. *Focus on Exceptionality*, *42*(8), 1–14.

Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions*, *11*, 133–144. doi:10.1177/1098300709332067

Horner, R. H., Todd, A. W., Lewis-Palmer, T., Irvin, L. K., Sugai, G., & Boland, J. B. (2004). The School-Wide Evaluation Tool (SET): A research instrument for assessing school-wide positive behavior support. *Journal of Positive Behavior Interventions*, *6*, 3–12. doi:10.1177/10983007040060010201

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. doi.10.1080/10705519909540118

Kincaid, D., Childs, K., & George, H. P. (2010). School-wide Benchmarks of Quality (BoQ). Unpublished instrument. Department of Child and Family Studies, University of South Florida, Tampa. Available from http://flPBIS.fmhi.usf.edu

Kincaid, D., Childs, K., Wallace, F., & Blase, K. (2007). Identifying barriers and facilitators in implementing schoolwide positive behavior support. *Journal of Positive Behavior Interventions*, *9*, 178–188.

McIntosh, K., Mercer, S. H., Hume, A. E., Frank, J. L., Turri, M. G., & Mathews, S. (2013). Factors related to sustained implementation of schoolwide positive behavior support. *Exceptional Children*, *79*, 293–311. doi:10.1177/1098300712470723

Mowbray, C., Holter, M., Teague, G., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, *24*, 315–340.

Newcomer, L. (2009). Universal positive behavior support for the classroom. *PBIS Newsletter*, *1*(4). Available from www.pbis.org

Pas, E. T., & Bradshaw, C. P. (2012). Examining the association between implementation and outcomes. *Journal of Behavioral Health Services & Research*, *39*, 417–433. doi:10.1007/s11414-012-9290-2

Pas, E. T., Waasdorp, T. E., & Bradshaw, C. P. (2014). Examining contextual influences on classroom-based implementation of positive behavior support strategies: Findings from a randomized controlled effectiveness trial. *Prevention Science*. Advance online publication. doi:10.1007/s11121-014-0492-0

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling*. Thousand Oaks, CA: SAGE.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: SAGE.

R Development Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. Retrieved from http://www.jstatsoft.org/v48/i02/

Sadler, C., & Sugai, G. (2009). Effective behavior and instructional support: A district model for early identification and prevention of reading and behavior problems. *Journal of Positive Behavior Interventions*, *11*, 35–46.

Simonsen, B., Eber, L., Black, A., Sugai, G., Lewandowski, H., Myers, D., & Sims, B. (2012). Positive behavioral interventions and supports in Illinois: Lessons learned for large-scale implementation. *Journal of Positive Behavior Interventions*, *14*, 5–16. doi:10.1177/1098300711412601

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education & Treatment of Children*, *31*, 351–380. doi:10.1353/etc.0.0007

Sugai, G., & Horner, R. H. (2006). A promising approach for expanding and sustaining school-wide positive behavior support. *School Psychology Review*, *35*, 245–259.

Sugai, G., & Simonsen, B. (2012, June 19). *Positive behavioral interventions and supports: History, defining features and misconceptions*. Retrieved from www.pbis.org

Tobin, T., Vincent, C., Horner, R., Rossetto, C., Dickey, C., & May, S. (2012). Fidelity measures to improve implementation of positive behavioural support. *International Journal of Positive Behavioural Support*, *2*(2), 12–19.

Vincent, C., Spaulding, S., & Tobin, T. J. (2010). A reexamination of the psychometric properties of the School-Wide Evaluation Tool (SET). *Journal of Positive Behavior Interventions*, *12*, 161–179.

Waasdorp, T. E., Bradshaw, C. P., & Leaf, P. J. (2012). The impact of schoolwide positive behavioral interventions and supports on bullying and peer rejection: A randomized controlled effectiveness trial. *Archives of Pediatrics & Adolescent Medicine*, *166*, 149–156. doi:10.1001/archpediatrics.2011.755