Hammill Institute on Disabilities

Behavioral Disorders
2018, Vol. 43(3) 357–369

DOI: 10.1177/0198742917747589
journals.sagepub.com/home/bhd
SAGE

Longitudinal Associations Between SWPBIS Fidelity of Implementation and Behavior and Academic Outcomes

Jerin Kim, MS[1], Kent McIntosh, PhD[1], Sterett H. Mercer, PhD[2], and Rhonda N. T. Nese, PhD[1]

Abstract

The purpose of this study was to examine associations between implementation fidelity of school-wide positive behavioral interventions and supports (SWPBIS) and student outcomes over time. The sample included 477 K–12 schools across 10 states implementing SWPBIS for varying numbers of years and at varying levels of fidelity (but with 78% of schools at or above established fidelity criteria). Results indicate that, in general, schools showed a decline in office discipline referrals (ODRs) and out-of-school suspensions (OSSs) over 3 years. SWPBIS fidelity of implementation was positively associated with initial levels of ODRs and OSSs, and the relations between fidelity and OSSs varied based on years of SWPBIS implementation. Levels of fidelity did not predict change in ODRs or OSSs. Despite no statistically significant associations between fidelity and student academic outcomes, schools that had implemented SWPBIS for 3 years or more had higher achievement in mathematics after controlling for prior achievement.

Keywords

SWPBIS, school-wide positive behavioral interventions and support, student outcomes

School-wide positive behavioral interventions and supports (SWPBIS) is a multitiered preventive behavior support framework that aims to reduce problem behaviors and improve learning environments (Horner et al., 2009). Schools implementing SWPBIS deliver a continuum of tiered support to students, including universal supports for all students, secondary supports for students at risk for challenges, and tertiary support for individual students in need of highly intensive supports (Sugai & Horner, 2006). Within universal supports, essential components include defining and teaching expected behaviors, encouraging prosocial behaviors, discouraging problem behaviors, making decisions with data, and implementing through a team leadership process (Sugai & Horner, 2006). It is important to maximize efforts at the universal level because well-established universal systems allow more students to be successful and reduce the number of students requiring support at the secondary and tertiary tiers (Kim, McIntosh, & Hoselton, 2014).

Multiple randomized controlled effectiveness trials have documented a range of positive effects of SWPBIS, such as reduced problem behaviors (Bradshaw, Mitchell, & Leaf, 2010; Horner et al., 2009; Pas, Waasdorp, & Bradshaw, 2015). These trials included at least a 2-year period of specific training and on-site coaching for school teams to implement SWPBIS with adequate fidelity of implementation—or the degree to which the intervention is delivered as designed, with accuracy and consistency (O'Donnell, 2008). In these trials, trainings produced adequate fidelity of implementation in treatment schools, and SWPBIS produced positive student outcomes. A conservative view of the findings from these trials is that the outcomes demonstrated can be expected not simply from adoption and training in SWPBIS but only if schools implement it with fidelity (Blase & Fixsen, 2013).

The Importance of Fidelity of Implementation

Fidelity of implementation relates to the extent to which core components of an intervention are implemented as intended by the developers (Century, Rudnick, & Freeman, 2010). Thus, specification of fidelity requires an

[1]University of Oregon, Eugene, USA
[2]University of British Columbia, Vancouver, USA

Corresponding Author:
Kent McIntosh, College of Education, University of Oregon, 141 Lokey Education Building, Eugene, OR 97403-1235, USA.
Email: kentm@uoregon.edu

understanding of varying aspects of the operational features (or components) of the originally designed intervention, such as intervention delivery, competence of intervention providers, and/or expected responsiveness of participants (Century et al., 2010; Dane & Schneider, 1998). Multiple measures, such as the School-Wide Evaluation Tool (SET; Sugai, Lewis-Palmer, Todd, & Horner, 2001) and School-Wide Benchmarks of Quality (BoQ; Cohen, Kincaid, & Childs, 2007), have been developed to assess the fidelity of implementation of SWPBIS. Total and subscale-level percentage scores are commonly used to indicate the proportion of the enacted components of SWPBIS. Higher scores indicate greater fidelity, and it is recommended that scores meet or exceed a criterion (e.g., 80% implementation on the SET) to ensure adequate implementation of enough critical features to improve student outcomes. These fidelity instruments serve multiple purposes (Algozzine et al., 2010). In addition to research-driven use, leadership teams also use fidelity measures to guide their ongoing implementation efforts toward full implementation. Also, state- or district-level coordinators and coaches can use them to assess schools' implementation and identify needs for improvement in their systems.

Associations Between Fidelity of Implementation and Student Outcomes

Fidelity of implementation is one variable of interest in many effectiveness trials because the outcomes of evidence-based practice could vary as a function of the quality of implementation (George & Childs, 2012). Based on this assumption, a growing number of researchers have investigated the associations between fidelity and student outcomes from large-scale samples of schools implementing SWPBIS with a wide range of fidelity.

Effects of Fidelity on Behavior Outcomes

Childs, Kincaid, George, and Gage (2016) evaluated student discipline outcomes over 4 years (2010–2011 to 2013–2014) from a sample of 1,122 Florida schools (724 elementary schools, 248 middle schools, and 150 high schools). In addition to a decrease in disciplinary exclusions (office discipline referrals [ODRs], in-school suspensions [ISSs], and out-of-school suspensions [OSSs]) across time, they found that higher BoQ scores (as measured by average scores across the 4 years) were significantly related to lower initial levels (intercepts) of disciplinary exclusions. However, higher BoQ total scores were not related to significantly lower or higher growth rates in ODRs and OSSs, and were modestly related to a higher growth rate (slope) of ISSs. Childs et al. concluded that schools with high fidelity had fewer disciplinary exclusions than those with low fidelity, whereas high fidelity did not produce faster reductions in disciplinary exclusions, suggesting that schools with higher fidelity experienced an immediate drop in exclusions but perhaps not a sustained steep drop that continued for multiple years.

In addition, Simonsen et al. (2012) investigated the association between fidelity of implementation and student outcomes from a sample of 428 Illinois schools (274 elementary/K–6th, 46 elementary/K–8th, 91 middle/6th–8th, and 17 high/9th–12th schools) implementing SWPBIS from 2000 to 2008. Throughout the study, the proportion of schools implementing SWPBIS with SET scores equal to or greater than the criterion increased from 36% to 78%, and schools showed significantly decreasing ODRs and mostly stable OSSs and total suspensions over 7 consecutive years. From 242 schools with both fidelity and suspension data, schools meeting or exceeding the SET criterion (as a time-variant covariate) had significantly fewer OSSs and total suspensions across years. Similarly, from 400 schools with fidelity and ODR data, implementation with fidelity was related to marginally lower ODRs per 100 students across years.

Freeman et al. (2016) investigated the association between implementation of SWPBIS with fidelity and student outcomes from a sample of 883 comprehensive high schools across 37 states that reported fidelity data at least once within a 7-year period. The main effects of fidelity on problem behavior were statistically significant and negative over multiple years. In comparison with schools with weak fidelity (e.g., equal to or less than 40% fidelity of implementation, as measured by the SET), ODR rates were significantly lower for schools with medium levels of fidelity (e.g., SET scores between 40% and 80%), and much lower for those with strong fidelity (e.g., SET scores equal to or greater than 80%). In addition, Flannery, Fenning, Kato, and McIntosh (2014) conducted a multi-level longitudinal analysis of individual behavior outcomes from 36,653 students (M = 1,770 per school) of 12 high schools (including four comparison schools) in two states over 3 years. In addition to significant effects of SWPBIS implementation on reductions in ODRs, there was a significant negative association between SET scores and ODRs in both the second and third years of the study.

Effects of Fidelity on Academic Outcomes

The research efforts to link fidelity and student outcomes have extended to academic achievement based on the assumption that reductions in problem behaviors due to implementation of SWPBIS with fidelity create a favorable learning environment, leading to improved academic achievement. Pas and Bradshaw (2012) investigated the association between fidelity of implementation (in 2008–2009) and student outcomes (in 2009–2010) with a sample

of 421 elementary and middle schools implementing SWPBIS after training and reporting data on at least one measure of fidelity in Maryland. Multiple fidelity measures were used, including the SET, BoQ, and the Implementation Phases Inventory (IPI; Bradshaw, Debnam, Koth, & Leaf, 2009), another SWPBIS fidelity of implementation measure. IPI scores predicted increased mathematics and reading achievement and decreased truancy, whereas no significant association between IPI scores and suspensions was found. Also, there were no significant associations between either BoQ or SET and all evaluated outcomes.

Simonsen et al. (2012) found in an 8-year study of reading and mathematics achievement data from 324 schools with fidelity and achievement data that schools meeting the 80% SET criterion showed a significantly higher proportion of students meeting or exceeding grade-level benchmarks on the statewide mathematics achievement test than their counterparts, whereas significant differences in the reading achievement test were not found. In addition, targeting high schools, Freeman et al. (2016) found no significant associations between fidelity and academic achievement (as a latent variable estimated by reading, language, and mathematics indices). Recently, Gage, Leite, Childs, and Kincaid (2017) examined the effects of implementation of SWPBIS with fidelity on academic achievement over 10 years from a sample of all elementary schools in Florida. Through two-level mixed effect regression models, they found implementation of SWPBIS at criterion (on the BoQ) was significantly related to higher proportions of students meeting or exceeding grade benchmarks in both reading and mathematics, yet the effects were weak. Although the main effects of fidelity scores were not evaluated, their findings suggest positive and distal effects of fidelity of SWPBIS implementation on academic achievement.

Summary

Taken together, research offers preliminary evidence suggesting a potentially positive association between fidelity of SWPBIS and disciplinary exclusions and academic achievement. Specifically, two studies (Gage et al., 2017; Pas & Bradshaw, 2012) suggested benefits of implementation of SWPBIS with fidelity for academic achievement, with relatively more evidence in mathematics (Simonsen et al., 2012). Yet, there are too few studies to determine convincingly whether SWPBIS improves academic outcomes, which indicates the need for further examination of the association between fidelity and academic achievement. In addition, the previous studies (Childs et al., 2016; Flannery et al., 2014; Freeman et al., 2016; Simonsen et al., 2012) found that schools with higher fidelity scores were more likely to have lower levels of disciplinary exclusions. By using the average of the BoQ scores across 4 years as a time-invariant covariate, the findings of Childs et al. (2016) that higher BoQ scores produced an immediate initial drop in disciplinary outcomes but did not affect the rates in later years suggest that student outcomes may be more responsive to varying levels of fidelity during the initial period of implementation than during later years. However, only a few studies (e.g., Childs et al., 2016) have examined the association between fidelity and growth rate of behavior outcomes. These analyses either treated years implementing as a nuisance variable (Childs et al., 2016; Simonsen et al., 2012) or assumed that those effects are similar (even when fidelity was handled as time-variant covariate; Simonsen et al., 2012), which does not allow assessment of growth rates and any systematic variations in growth over time.

Despite a lack of empirical evidence for the number of years that comprise initial implementation (and thus might show the strongest effects of SWPBIS), previous trials (Bradshaw et al., 2010; Flannery et al., 2014; Horner et al., 2009) assumed this period may span at least 2 years. As such, it is unknown whether these effects are seen only for schools initially implementing SWPBIS or if these findings are consistent past the 2-year implementation window assessed in previous trials. Thus, additional research is needed to examine the association between fidelity and growth trajectories of student outcomes, and assess whether the effects of fidelity on student outcomes each year during the initial years of implementation are different from those that have been implementing for more years.

Purpose of the Study

With an emphasis on the association between SWPBIS fidelity of implementation and student outcomes, it is important to examine the extent to which fidelity is associated with improved outcomes as schools move from initial implementation to sustained implementation (McIntosh et al., 2014). In this regard, this article examined the effects of fidelity on initial levels and growth rates of student outcomes. Specifically, this study aimed to answer the following questions:

Research Question 1: To what extent is SWPBIS fidelity of implementation related to initial level and change in school-level behavior outcomes (ODRs and OSSs), and do effects vary by years implementing SWPBIS?
Research Question 2: To what extent is SWPBIS fidelity of implementation related to school-level academic outcomes (reading and mathematics achievement), and do effects vary by years implementing SWPBIS?

Method

Participants and Settings

The sample included 477 schools across 10 U.S. states. These schools were part of a larger project assessing

Table 1. School Demographics.

Characteristics	*n*	%
Years implementing SWPBIS[a]		
0–2 years	185	38.78
3 or more years	292	61.22
School level		
Elementary schools	342	71.69
Middle schools	92	19.29
High schools	41	8.59
School locale		
Rural	89	18.66
Town	77	16.14
Suburb	152	31.87
City	159	33.33
M % of minority students (*SD*)		43.85 (30.59)
M % of students eligible for FRL		53.63 (23.92)

Note. SWPBIS = school-wide positive behavioral interventions and supports; FRL = free or reduced-price lunch.
[a]The average years of implementation is 3.65 (*SD* = 2.69); the total sample size is 477.

longitudinal implementation of SWPBIS (McIntosh et al., 2017). The sample for the current study included all regular schools with discipline data available from the School-Wide Information System (SWIS), a web-based data management system for implementation of SWPBIS. All schools had complete demographic data from the National Center for Educational Statistics (NCES). Demographic data are reported in Table 1. During the first year of the study (2012–2013), schools had been implementing Tier I SWPBIS for an average of 3.65 years (*SD* = 2.69).

Measures

Behavior outcomes: ODRs per 100 students per school day. ODRs are administered to students for a specified set of behavior violations. ODR data were obtained from SWIS (www.pbisapps.org). For this study, only major ODRs, issued for more severe violations that require school administrator involvement, were included in counts. The total ODR counts during 1 school year were divided by school enrollment and number of school days (entered by schools but replaced by NCES data if missing) and multiplied by 100 in the year to produce the number of ODRs per 100 students per school day. On average, the ODRs per 100 students per school day were 0.32 (*SD* = 0.67) in 2012–2013, 0.29 (*SD* = 0.39) in 2013–2014, and 0.27 (*SD* = 0.35) in 2014–2015. To reduce strong positive skew of the data, we applied square root transformations before data analyses. All schools except one had complete ODR data for all three time points.

Behavior outcomes: OSSs per 100 students per school day. OSS data were also obtained from SWIS. An OSS is a complete exclusion from the school for a specified period (more restrictive than ISS, which excludes the student from the classroom but not the school). OSS values indicate the number of OSS events per 100 students per school day and were obtained by dividing the total count of OSSs by school enrollment and school days and multiplying by 100. On average, OSSs per 100 students per school day was 0.04 (*SD* = 0.07) in 2012–2013, 0.03 (*SD* = 0.06) in 2013–2014, and 0.03 (*SD* = 0.06) in 2014–2015. We also applied square root transformations to address skewness before analyses.

Academic outcomes (reading and mathematics achievement). We used school-level data reporting the proportion of students meeting or exceeding grade-level proficiency criteria for state reading and mathematics assessments, respectively, in 2012–2013 and 2014–2015. There were 464 (97.3%) and 469 (98.3%) schools with complete academic data in 2012–2013 and 2014–2015. Missing data occurred because schools had missing data or the state did not report data to the public for various reasons (e.g., new schools, small-size schools, achievement measure change). Due to variations in the achievement tests administered and relevant mastery criteria across states and over time within states, school data were state mean-centered and divided by state standard deviations each year to generate standardized scores as an indicator of within-state relative status of school performance.

SWPBIS fidelity of implementation. Schools administered one or more measures to assess fidelity of implementation of SWPBIS at Tier 1. Those research-validated measures include (a) the SET (Sugai, Lewis-Palmer, et al., 2001), (b) the BOQ (Kincaid, Childs, & George, 2005), (c) the SWPBIS Self-Assessment Survey (SAS; Sugai, Horner, & Todd, 2000), and (d) the Team Implementation Checklist (TIC; Sugai, Todd, & Horner, 2001). Total percentage scores (obtained by dividing the sum scores by the available maximum scores) in 2012–2013 were used for this study. In 2012–2013, SET scores were reported by 271 schools (*M* = 91%), SAS scores were reported by 293 (*M* = 78%), TIC scores were reported by 176 schools (*M* = 74%), and BOQ scores were reported by 228 schools (*M* = 83%). Overall, 78.2% of fidelity assessments were at or above the measure's criterion, indicating that the vast majority of schools were implementing SWPBIS with adequate fidelity. Because schools in the study did not use a consistent fidelity measure, we followed the procedures of Turri and colleagues (2016) to create a single latent variable. This continuous latent variable, SWPBIS Fidelity, included all available observed scores from the SET, BOQ, SAS, and TIC in 2012–2013, the first year of the study, as indicators.

This variable had a composite reliability of .90. Details of each instrument are described below.

SET. The SET (Sugai, Lewis-Palmer, et al., 2001) is a research-validated measure of fidelity of implementation of SWPBIS at the Tier 1 that was mainly developed as a research tool. The SET involves 28 research questions in seven subscales: Expectations Defined, Expectations Taught, On-going System for Rewarding Expectations, System for Responding to Violations, Monitoring & Decision-Making, Management, and District-Level Support. Each item is scored by an external coach using the Scoring Guide that determines specific criteria for 0 to 2 point scores. Research (Horner et al., 2004) revealed strong internal consistency (α = .96), test–retest reliability (*mean agreement* = 97%), interrater reliability (*mean agreement* = 99%), and concurrent validity (correlation with SAS = .75).

School-wide BoQ. The BoQ (Kincaid et al., 2005) is a reliable and valid measure of fidelity of implementation of SWPBIS at the Tier 1. This tool includes 53 items and 10 subscales: SWPBIS Team, Faculty Commitment, Effective Procedures for Dealing With Discipline, Data Entry and Analysis Plan Established, Expectations and Rules Developed, Reward/Recognition Program Established, Lesson Plans for Teaching Expectations/Rules, Implementation Plan, Classroom Systems, and Evaluation. The BoQ was developed as a self-rating tool for school teams and their external coach to rate each item together using the scoring rubric. Depending on significance, individual items have varying maximum point scores (1–3). The details of the instrument provide information about the implementation process and guide leadership team's effective and efficient action planning. The BoQ has been proven to have strong internal consistency (α = .96), test–retest reliability (r = 94%), and interrater reliability (r = 87%) as well as moderate correlations (r = .51) with the SET (Cohen et al., 2007).

PBIS SAS. The SAS (Sugai et al., 2000) intends to assess school staff's perception of the implementation status and priority for improvement. This instrument is used as an initial or annual assessment of four behavior support systems: school-wide systems (18 items), nonclassroom systems (nine items), classroom systems (11 items), and individual student support systems (eight items). Particularly, the rating of the current status of school-wide system indicates the fidelity of implementation for Tier 1 of SWPBIS. Research (Hagan-Burke et al., 2005; Horner et al., 2004; Solomon, Tobin, & Schutte, 2015) indicates that the SAS has adequate internal consistency (e.g., α = .88; Hagan-Burke et al., 2005) and convergent validity (e.g., correlation with SET r = .75, Horner et al., 2004).

TIC. The TIC (Sugai, Todd, et al., 2001) is used as a self-rating tool for effective Tier 1 implementation. To monitor implementation progress, school teams regularly administer this instrument by scoring via 3-point (0–2) Likert-type scales. The TIC includes 22 items across seven subscales: Establish Commitment, Establish and Maintain Team, Self-Assessment, Establish School-Wide Expectations: Prevention Systems, Classroom Behavior Support Systems, Establish Information Systems, and Build Capacity for Function-Based Support. The prior research indicates strong reliability (e.g., ordinal α = .95) and factorial validity (via a confirmatory factorial analysis; McIntosh, Mercer, Nese, Strickland-Cohen, & Hoselton, 2016; Mercer, McIntosh, & Hoselton, 2017).

Contextual variables. School demographic data from 2012–2013 were obtained from the NCES (2014). These variables included school locale (a categorical variable of rural, town, suburban, or urban), the percentage of students eligible for free or reduced-price lunch (FRL), the percentage of non-White students enrolled in the school, and grade levels served (1 = *elementary*, 0 = *non-elementary*). In addition, schools were divided by years of implementing SWPBIS into two groups: schools implementing Tier I SWPBIS for 0 to 2 years (Initial Implementers, n = 185) or 3 or more years (Sustained Implementers, n = 292).

Procedure

Within a larger longitudinal research project sample, schools using SWIS to enter and analyze school discipline data were selected for this study. With school consent to use data for research in accordance with an approved institutional review board (IRB) protocol, we obtained SWIS ODR and OSS data and fidelity of implementation scores from 2012–2013 to 2014–2015 from a database of discipline data and fidelity of implementation maintained by the OSEP National Technical Assistance Center on PBIS. We collected academic proficiency (measured by percentage of students reported as Proficiency or above) in Reading (or English/Language Arts) and Mathematics from publicly available state aggregate files (e.g., http://www.ode.state.or.us/apps/BulkDownload/BulkDownload.Web/) or individual school report cards (e.g., http://www.ode.state.or.us/data/reportcard/reports.aspx). The school-level reading and mathematics proficiency percentages reported in School Report Cards were selected if available, and if needed, grade-level proficiency percentages were averaged to generate the school-level aggregate score. In addition to periodic accuracy checks during data entry, 10% of schools were randomly selected at the conclusion of data entry—all values for the selected schools were verified as correct.

Data Analysis

To examine the associations between fidelity and student behavior outcomes over 3 years, we conducted growth linear modeling (GLM) separately for the two behavior outcomes

using M*plus* 7 (Muthén & Muthén, 1998–2012). We modeled the continuous fidelity factor score in 2012–2013 as the primary predictor of intercept and slope to evaluate the associations between fidelity of implementation and initial levels and change in behavior outcomes across 3 years. In specific, we examined whether schools implementing SWPBIS with higher levels of fidelity in 2012–2013 had (a) lower levels of disciplinary exclusions in 2012–2013 and (b) more rapid improvement in following years. We also included an interaction effect of fidelity by years implementing SWPBIS (also including the main effect of years implementing) to examine whether fidelity effects on disciplinary exclusions were consistent between initial implementer schools (implementing SWPBIS for 2 or less years) and sustained implementer schools (implementing SWPBIS for 3 or longer years). To model the interaction term between the continuous latent variable (SWPBIS Fidelity) and dichotomous observed variable (years implementing SWPBIS; 3 or more years vs. 2 or fewer years), the XWITH (short for multiplied with) command was used under TYPE = RANDOM (used for the random slopes) and ALGORITHM = INTEGRATION (used for the numerical integration with, for example, the model with interaction terms involving the latent variables of which posterior distribution has no closed form expression; Muthén & Muthén, 2009). Then, we entered school contextual variables as time-invariant covariates and only retained significant predictors in final models, aligned with all previous large-scale studies controlling for multiple demographic variables (e.g., grade levels, enrollment, proportion of students receiving FRL).

As a visual aid to help interpret the findings, ODR and OSS patterns of change over 3 years were plotted depending on fidelity of implementation for two different groups of schools, the sustained and initial implementer schools. To create the dichotomous fidelity variable (for easier interpretation), continuous fidelity factor scores were obtained from the measurement model (specifying the fidelity factor as observed by BoQ, TIC, SAS, and SET total percentage scores), and recoded into a dichotomous variable indicating 1 = *higher fidelity* (greater than the mean) and 0 = *lower fidelity* (equal to or lower than the mean). Then, a total of four groups of schools depending on fidelity and years of implementation (e.g., initial implementer schools with higher fidelity) were graphed and compared for each outcome.

For academic achievement, state variations in the tests (e.g., shifting to and from measures aligned with the Common Core State Standards) created instability in scores that limited the use of growth modeling, and therefore, regression models were used to examine whether school differences in fidelity in 2012–2013 relate to the relative within-state academic performance of schools in 2014–2015 after controlling for their academic performance in 2012–2013. SWPBIS Fidelity in 2012–2013 was modeled as a primary predictor. Years implementing and an interaction term (SWPBIS Fidelity by years implementing) were also included in models. Other contextual variables were entered, and only significant variables were retained for the final analyses.

For all models, model fit was evaluated using multiple fit indices and criteria (Hu & Bentler, 1999; Steiger, 2007): chi-square, $p > .05$, comparative fit index (CFI) $\geq .95$, root mean square error of approximation (RMSEA) < 0.06, and its confidence interval (upper limit) < 0.07. However, when using numerical integration to estimate latent interaction term, chi-square and other fit statistics are not calculated; therefore, Akaike information criterion (AIC) and Bayesian information criteria (BIC) were used to evaluate models with interaction terms for behavior outcomes. The nested data structure (schools within districts) was addressed in the model by standard error adjustment via the TYPE = COMPLEX command (to treat district as a nuisance variable) in M*plus*. From the dataset, only 13 schools (2.7%) had any missing data in student behavior or academic outcomes for all years assessed, and 47 schools (9.85%) did not have fidelity data in 2012–2013. These 47 schools without fidelity data in 2012–2013 were included because they had fidelity data in subsequent years. Using listwise deletion to remove these schools from analyses has been shown to bias results more than using newer missing data handling approaches, such as maximum likelihood (ML) estimation, as used here (Enders, 2010). However, ML estimation is not applicable to observed exogeneous data, and eight schools lacking reading and mathematics achievement data in 2012–2013 were excluded from academic models. All results of each model were reported via using the unstandardized coefficients because standardized coefficients were not available in TYPE = RANDOM estimation.

Results

Tables 2 and 3 present the results of the baseline and/or final models for each outcome. All final models exhibited adequate model fit based on the predetermined criteria, except for a statistically significant chi-square test for the mathematics model. Prior to estimating the conditional models for behavior outcomes, unconditional models were examined to evaluate the intercept and slope without covariates. As indicated by Table 2, the statistically significant negative slopes observed for behavior outcomes indicated that overall, means for both ODRs and OSSs decreased considerably over the 3-year time period for schools implementing SWPBIS. Figures 1 and 2 provide visual representations of ODR and OSS means over time by SWPBIS Fidelity (above the mean vs. below the mean) and years implementing (0 to 2 vs. 3 or more years).

Table 2. Associations Between SWPBIS Fidelity and Student Behavior.

	ODRs		OSSs	
Variables	Unconditional	Conditional	Unconditional	Conditional
Predictors of intercept				
SWPBIS fidelity		0.724*		0.228*
		(0.331)		(0.111)
Years implementing SWPBIS		−0.016		0.013
		(0.037)		(0.015)
Fidelity by years		−0.668		−0.276*
		(0.366)		(0.127)
% of students eligible for FRL		0.354***		0.184***
		(0.087)		(0.037)
% of minority students		−0.382***		
		(0.072)		
Elementary school		−0.078**		−0.081***
		(0.030)		(0.011)
Predictors of slope				
SWPBIS fidelity		−0.356		−0.072
		(0.257)		(0.064)
Years implementing SWPBIS		−0.007		−0.012
		(0.028)		(0.009)
Fidelity by years		0.373		0.161
		(0.274)		(0.088)
% of minority students		0.083*		
		(0.042)		
Means				
Intercept	0.457***	0.510***	0.141***	0.097***
	(0.030)	(0.063)	(0.014)	(0.019)
Slope	−0.027*	−0.065**	−0.015**	−0.010
	(0.014)	(0.023)	(0.005)	(0.007)
Model fit indices				
Chi-square	0.601		0.388	
df	2		3	
Comparative fit index	1.000		1.000	
Root mean square error of approximation (value)	0.000		0.000	
Root mean square error of approximation (upper limit)	0.063		0.014	
Akaike information criterion	484.541	−271.149	−2,343.156	−3,298.359
Bayesian information criterion	513.713	−138.131	−2,443.151	−3,177.501

Note. Standard errors of coefficients were reported within parenthesis. SWPBIS = school-wide positive behavioral interventions and supports; ODRs = office discipline referrals; OSSs = out-of-school suspensions; FRL = free or reduced-price lunch.
*$p < .05$. **$p < .01$. ***$p < .001$.

ODRs

Results showed that SWPBIS Fidelity was positively related to the ODR intercept (higher fidelity in 2012–2013 predicting higher ODRs in 2012–2013; $\beta = .724$), but years implementing and the fidelity by years implementing interaction were not significantly related to the ODR intercept. Neither SWPBIS Fidelity nor years implementing were significant predictors of slope, although both terms were negative (higher fidelity in 2012–2013, slightly slower growth in ODRs).

Figure 1 displays these patterns and the interaction effect. It shows that the highest initial ODR rates were seen for sustained implementer schools with fidelity below the mean. All groups showed a decreasing trend in ODR rates except for initial implementer schools with fidelity below the mean. Instead of decreases, these schools had a fluctuating pattern over time, with a sharp increase in ODRs in the second year, remaining slightly above the earlier mean level in the third year. In 2014–2015, sustained implementer schools saw lower ODR rates than initial implementer schools.

Table 3. Associations Between SWPBIS Fidelity and Student Academic Outcomes.

Variables	Reading	Mathematics
SWPBIS Fidelity	0.259	0.374
	(0.369)	(0.418)
Years implementing SWPBIS	0.096	0.438***
	(0.077)	(0.064)
Fidelity by years	0.463	−0.219
	(0.539)	(0.580)
% of students receiving FRL	−1.168***	−1.685***
	(0.236)	(0.257)
Reading/mathematics in 2012–2013	0.611***	0.438***
	(0.062)	(0.064)
Model fit indices		
Akaike information criterion	99.822	150.288
Bayesian information criterion	191.135	245.752

Note. Standard errors of coefficients were reported within parenthesis. SWPBIS = school-wide positive behavioral interventions and supports; FRL = free or reduced-price lunch; reading = proportion of students meeting or exceeding state-level standards in reading achievement; mathematics = proportion of students meeting or exceeding state-level standards in mathematics achievement.
*$p < .05$. **$p < .01$. ***$p < .001$.

The significant contextual predictors of ODR intercept were percentage of students eligible for FRL (predicting higher ODRs; β = .354), percentage of non-White students (predicting lower ODRs; β = –.382), and elementary school (predicting lower ODRs, β = –.078), whereas the percentage of non-White students was only a significant predictor of the slope (predicting increasing ODRs; β = .083).

OSSs

Consistent with the results for ODRs, SWPBIS Fidelity was positively related to the OSS intercept (higher fidelity in 2012–2013 predicting higher OSSs in 2012–2013; β = .228) and years implementing was not significantly related to the OSS intercept, but a significant negative interaction effect of SWPBIS Fidelity by years implementing (β = –.276) showed that differences in levels of OSSs between initial and sustained implementer schools became smaller with increasing levels of fidelity. Neither SWPBIS Fidelity nor years implementing were significant predictors of OSS slope, although both terms were negative.

Like Figure 1, Figure 2 shows the highest OSS rates in 2012–2013 for sustained implementer schools with fidelity below the mean. As with ODRs, all groups showed decreasing OSSs over time, except for initial implementer schools with fidelity below the mean. These schools saw increased OSS rates in the second year, which then returned to initial levels at the third year.

Significant contextual predictors of OSS intercept were the percentage of students eligible for FRL (predicting higher OSSs; β = .184) and elementary schools (predicting lower OSSs; β = –.081). No contextual variables significantly predicted the OSS slope.

Academic Achievement

Results showed that SWPBIS Fidelity was a positive but not statistically significant predictor of both reading and mathematics achievement in 2014–2015, after controlling for prior achievement in 2012–2013. Years implementing was also a positive but not significant predictor of reading achievement. However, years implementing (β = .438) was significantly positively associated with mathematics achievement in 2014–2015, such that sustaining implementer schools had significantly higher achievement. The interaction effects of fidelity by years implementing were not significant in reading and mathematics achievement as well. For both outcomes, the percentage of students eligible for FRL significantly negatively predicted achievement in reading (predicting lower achievement; β = −1.168) and mathematics (predicting lower achievement; β = −1.685).

Discussion

This study examined the effects of fidelity of SWPBIS implementation on growth trajectories of behavior and academic outcomes for students. Descriptive results revealed that mean fidelity was high for the entire sample, and mean disciplinary removals (ODRs and OSS) decreased over the 3-year period. However, there were no significant associations between fidelity and change in behavior or academic outcomes, although sustaining implementer schools showed better performance in mathematic achievement. Although the findings related to behavior outcomes appear perplexing, it should be noted that all of the schools in the sample had been provided SWPBIS training, 78% of the schools were implementing SWPBIS with fidelity in 2012–2013, and both groups (i.e., the initial implementer schools and the sustaining implementer schools) were likely to have been implementing enough critical features of SWPBIS to improve student outcomes. Therefore, the study provides a test of an additive effect of high fidelity of implementation for schools that generally met fidelity of implementation criteria (i.e., comparing "high fidelity" with "at fidelity" implementation).

Consistent with the results of SWPBIS trials (Bradshaw et al., 2010; Flannery et al., 2014; Horner et al., 2009), there was a general decrease in ODRs and OSSs for schools implementing SWPBIS over the 3-year period for schools in the sample. However, the precise relation of SWPBIS fidelity to levels of disciplinary outcomes in our sample is a little less clear than in previous studies. For example, Flannery et al. (2014) reported that higher levels of fidelity of implementation (as a time-varying covariate) was

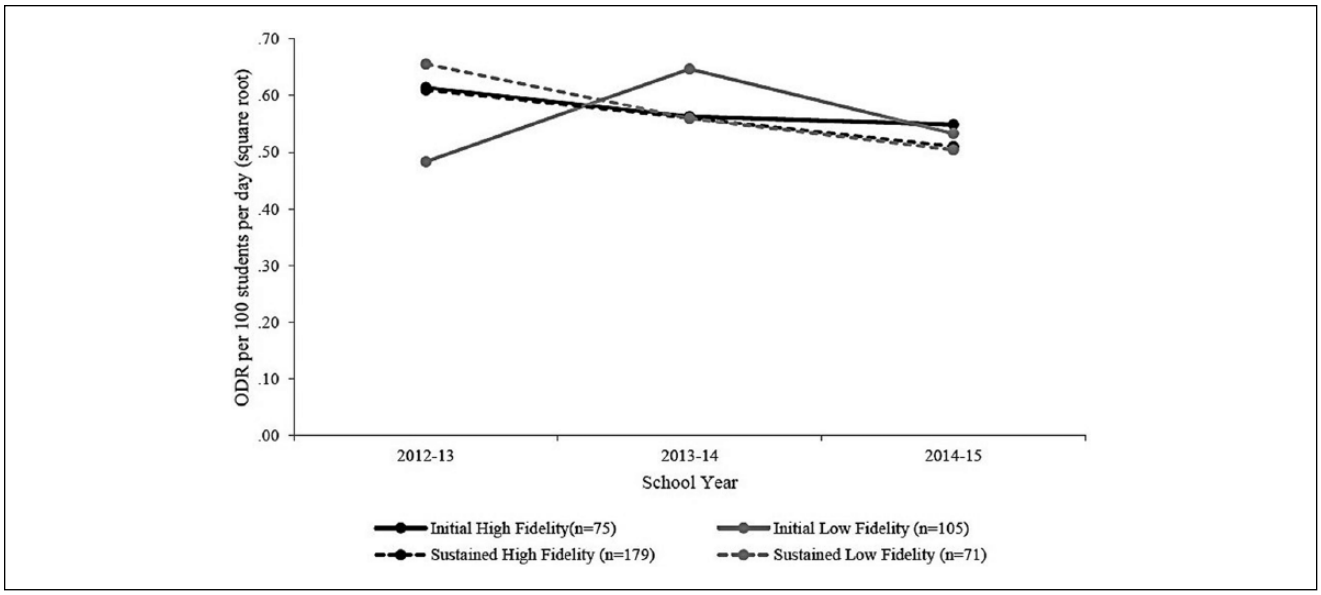

Figure 1. Mean differences in ODR over 3 years by SWPBIS Fidelity and year implementing SWPBIS.
Note. A total of 47 schools without fidelity data were excluded from this figure. ODR = office discipline referrals; SWPBIS = school-wide positive behavioral interventions and supports; initial implementers = 0 to 2 years of implementation: sustained implementers = 3 or more years of implementation; high fidelity = equal to or greater than mean fidelity factor score; low fidelity = lower than mean fidelity factor scores.

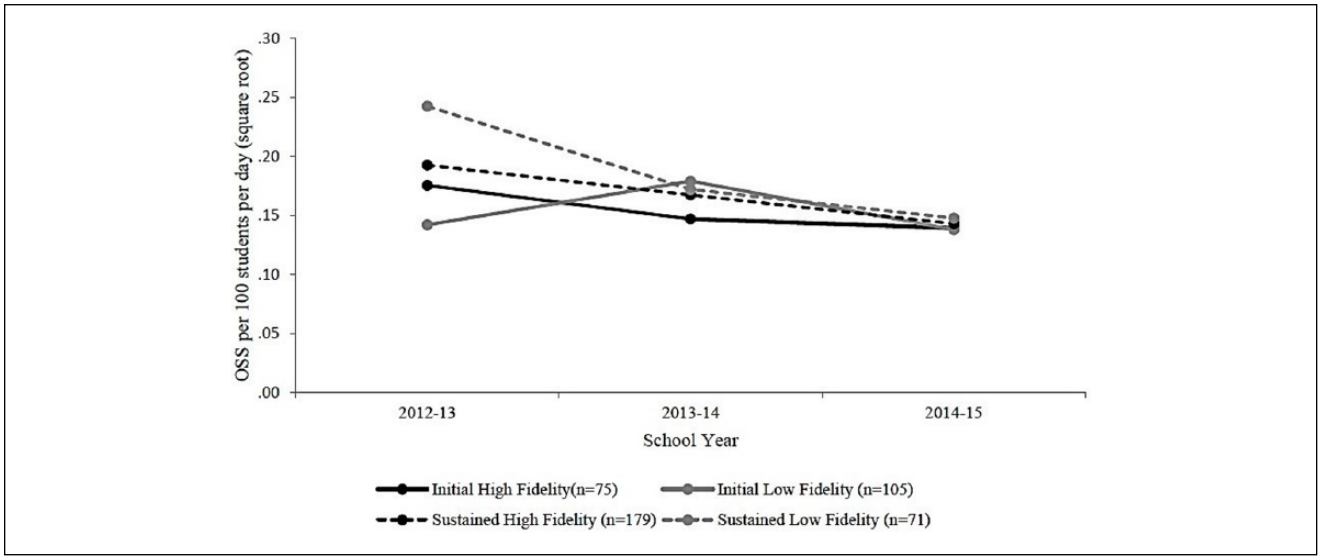

Figure 2. Mean differences in OSS over 3 years by SWPBIS Fidelity and years implementing SWPBIS.
Note. A total of 47 schools without fidelity data were excluded from this figure. OSS = out-of-school suspensions; SWPBIS = school-wide positive behavioral interventions and supports; initial implementers = 0 to 2 years of implementation; sustained implementers = 3 or more years of implementation; high fidelity = equal to or greater than mean fidelity factor score; low fidelity = lower than mean fidelity factor scores.

significantly related to lower levels of ODRs over the latter 2 years of implementing SWPBIS from the sample of high schools, which is consistent with other large-scale studies (Childs et al., 2016; Freeman et al., 2016; Simonsen et al., 2012). In contrast, the results of this study indicate a positive association between fidelity (in 2012–2013) and initial level of ODRs and OSSs. Regarding years implementing SWPBIS, the significant and negative interaction effect of fidelity by years implementing on the level of OSSs indicates that the positive effects of fidelity on the level of OSSs were stronger for initial implementer schools. In other words, the association between fidelity and level of OSSs was different (i.e., negative) for sustained implementer schools. These results contrasted with our expectation that

more obvious improvements in disciplinary exclusions would be made during the initial period (Childs et al., 2016), although the restricted range of fidelity in this sample should provide caution.

To understand the findings more precisely, it is helpful to consider the study sample. Figures 1 and 2 show that all groups displayed decreasing patterns in both ODRs and OSSs, except for initial implementer schools with fidelity below the mean. These initial implementer schools with lower fidelity had increased discipline exclusions in the second year, which then returned to initial levels at the third year. Also, these schools with lower fidelity had the lowest rates of ODRs and OSSs in 2012–2013. One hypothesis is that these initial implementer schools with lower fidelity were using their discipline data systems inconsistently, a core feature of Tier 1 SWPBIS, and therefore, issuing fewer ODRs and OSSs in their first year of implementation. Also, it could be that the increased ODR and OSS rates reflect more accurate measures of student behavior in the second year as schools moved forward with implementation, including defining and systematizing school discipline processes.

Also, this study found that the growth rates of OSSs or ODRs were not predicted by fidelity, which is consistent with Childs et al. (2016). One explanation may relate to whether SWPBIS is fully integrated into classroom practices and systems. Childs et al. (2016) examined the associations between subscale scores of BoQ and student outcomes, and found Classroom subscale scores were the only significant predictor of behavior outcomes, suggesting that adequate classroom-level implementation might be a critical feature in reducing discipline exclusions. Although stronger fidelity of school-wide implementation is often associated with stronger classroom-based implementation (Pas et al., 2015), schools meeting or exceeding the fidelity criteria for SWPBIS implementation may not have adequate implementation of positive behavior support systems in classrooms. For example, in their study of 33 elementary schools implementing SWPBIS with adequate fidelity, Reinke, Herman, and Stormont (2013) found that there were still needs for improvement in classroom mangement strategies (e.g., specific praise, 4:1 positive to negative interaction ratios) and documentation systems for rewarding desired behaviors and reporting unwanted behaviors. Accordingly, adequate fidelity of SWPBIS at the school level might not reduce unwanted behaviors, if classroom practices are not fully connected to the school-wide systems.

Regarding academic outcomes, years implementing was a stronger predictor of academic achievement than SWPBIS fidelity in this study. Results showed that both SWPBIS fidelity and years implementing were positive but not statistically significant predictors of reading achievement, but that sustaining implementer schools had significantly higher mathematics achievement. These findings may be indicative of the distal effects on student learning via improving behaviors as indicated by Gage et al. (2017). It is possible that under typical, less-controlled implementation, academic outcomes are more likely to improve after multiple years of implementing SWPBIS. Also, the benefit of years implementing SWPBIS was proven only in mathematics achievement, which might be related to the nature of the subject. One hypothesis is that variance in reading outcomes may be due more to variance in instructional methods (e.g., explicit vs. constructivist instruction), whereas instruction in mathematics may be more consistent, leading to more influence from other variables (Duncan et al., 2007).

Several demographic variables were included as covariates in this study. Proportion of students receiving FRL was the only covariate predicting the level of both (negatively) academic and (positively) behavior outcomes, which is consistent with prior studies (e.g., Flannery et al., 2014; Freeman et al., 2016; Gage et al., 2017). In relation to school levels, elementary schools had lower rates of ODRs and OSSs. These results appear to be reasonable considering structural barriers against systematic whole school implementation in secondary school settings, particularly high schools (Flannery et al., 2014).

Implications for Practice

The primary findings of the current research offer several implications for supporting effective implementation. First, it may be that implementing SWPBIS with fidelity is related to improved behavior and academic outcomes, but extremely high rates of fidelity do not appear to lead to significantly enhanced outcomes. Hence, technical assistance can focus on implementing the core features adequately. Second, the results showed that initial implementer schools displayed an unstable pattern of behavior outcomes in response to fidelity, indicating that this period of initial implementation may need more intensive technical assistance from the district or state, because improvement in behavior outcomes can reinforce continued implementation by school practitioners (Andreou, McIntosh, Ross, & Kahn, 2015). Third, the current study indicates that academic outcomes may not be realized until enough years have passed to establish effective academic learning environments and allow instruction to take place. Sharing this information may help evaluators and administrators have realistic expectations regarding academic outcomes and be aware that academic achievement is unlikely to improve in the absence of high-quality academic instruction (McIntosh & Goodman, 2016).

Limitations

One of the primary limitations of this study was the nature of the schools represented in the sample. The vast majority

of schools were implementing SWPBIS to criterion, including those in the "below the mean" groups in Figures 1 and 2. Therefore, the results more closely represent high versus very high implementation fidelity, rather than previous research, which has examined those implementing at or above criterion versus those below. This restricted range in the sample could explain how these findings contrast with those seen in previous research (Pas et al., 2015), and therefore should be interpreted with caution. Another limitation was the sources of academic achievement data used in this study. As a result of the adoption and de-adoption of Common Core State Standards at different years by some states, academic achievement tests changed variously across years and states. Due to the fluctuation in the tests used, the resulting student data limited the ability of assessing growth trajectories of academic outcomes as was done with behavior. In addition, the reliability of data entry of student discipline outcomes by schools (via SWIS) is unknown.

Implications for Future Research

This study proposed that there might be various patterns of student outcomes in a response to changing fidelity across different implementation stages. However, using only two time spans (0 to 2 vs. 3 or more years) might be too coarse to capture patterns in long-term implementation and change, which could require more granular measurement of years implementing. In recognition of the restricted range of fidelity in the study sample, more efforts can be made to obtain the natural and representative sample of schools, which needs the carefully designed sampling process. In terms of outcome measures, examining effects of typical SWPBIS implementation on additional variables not studied here (e.g., student–teacher relationships, school climate, racial disproportionality in disciplinary outcomes) would also be useful for stakeholders considering whether to adopt a SWPBIS initiative. Due to state variations in policy, standards, or curricula that may affect academic achievement, state-level predictors (e.g., state-level science, technology, engineering and mathematics [STEM] initiatives, funding, rigor of standards, teacher resources; Lee & Reeves, 2012) could be included in analytic models. Finally, analyzing changes in classroom context and delivery of instruction upon implementation of classroom SWPBIS systems may help to shed light on the precise mechanisms by which SWPBIS may impact and improve instruction. Improvement in instruction may be a result of a number of factors, including higher quality instructional practices, fewer distractions in the classroom as a result of students being more on-task, and feeling safer at school. As such, experimental research in these areas could identify the specific critical features of SWPBIS that are most predictive of improvements in both behavior and academic outcomes for students.

Authors' Note

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

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A120278 to University of Oregon.

ORCID iD

Kent McIntosh https://orcid.org/0000-0002-7765-2993
Sterett H. Mercer https://orcid.org/0000-0002-7940-4221

References

Algozzine, B., Horner, R. H., Sugai, G., Barrett, S., Dickey, C. R., Eber, L., . . . Tobin, T. (2010). *Evaluation blueprint for school-wide positive behavior support*. Eugene, OR: National Technical Assistance Center on Positive Behavior Interventions and Support. Retrieved from http://www.pbis.org/blueprintguidestools/blueprint/evaluation-blueprint

Andreou, T. E., McIntosh, K., Ross, S. W., & Kahn, J. D. (2015). Critical incidents in the sustainability of school-wide positive behavioral interventions and supports. *The Journal of Special Education*, *49*, 157–167. doi:10.1177/0022466914554298

Blase, K., & Fixsen, D. (2013). *Core intervention components: Identifying and operationalizing what makes programs work*. Washington, DC: Office of the Assistant Secretary for Planning and Evaluation, Office of Human Services Policy, U.S. Department of Health and Human Services. Retrieved from http://aspe.hhs.gov/hsp/13/KeyIssuesforChildrenYouth/CoreIntervention/rb_CoreIntervention.cfm

Bradshaw, C. P., Debnam, K., Koth, C. W., & Leaf, P. (2009). Preliminary validation of the Implementation Phases Inventory for assessing fidelity of schoolwide positive behavior supports. *Journal of Positive Behavior Interventions*, *11*, 145–160. doi:10.1177/1098300708319126

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 Behavior Interventions*, *12*, 133–148. doi:10.1177/1098300709334798

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, *31*, 199–218. doi:10.1177/1098214010366173

Childs, K. E., Kincaid, D., George, H. P., & Gage, N. A. (2016). The relationship between school-wide implementation of positive behavior intervention and supports and student discipline outcomes. *Journal of Positive Behavior Interventions*, *18*, 89–99. doi:10.1177/1098300715590398

Cohen, R., Kincaid, D., & Childs, K. E. (2007). Measuring school-wide positive behavior support implementation development and validation of the benchmarks of quality. *Journal of Positive Behavior Interventions*, *9*, 203–213. doi:10.1177/1098300707090040301

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*, 23–45. doi:10.1016/S0272-7358(97)00043-3

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Brooks-Gunn, J. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428–1446. doi:10.1037/0012-1649.43.6.1428

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Flannery, K. B., Fenning, P., Kato, M. M., & McIntosh, K. (2014). Effects of school-wide positive behavioral interventions and supports and fidelity of implementation on problem behavior in high schools. *School Psychology Quarterly*, *29*, 111–124. doi:10.1037/spq0000039

Freeman, J., Simonsen, B., McCoach, D. B., Sugai, G., Lombardi, A., & Horner, R. (2016). Relationship between school-wide positive behavior interventions and supports and academic, attendance, and behavior outcomes in high schools. *Journal of Positive Behavior Interventions*, *18*, 41–51. doi:10.1177/1098300715580992

Gage, N. A., Leite, W., Childs, K., & Kincaid, D. (2017). Average treatment effect of school-wide positive behavioral interventions and supports on school-level academic achievement in Florida. *Journal of Positive Behavior Interventions*, *19*, 158–167. doi:10.1177/1098300717693556

George, H. P., & Childs, K. E. (2012). Evaluating implementation of schoolwide behavior support: Are we doing it well? *Preventing School Failure: Alternative Education for Children and Youth*, *56*, 197–206.

Hagan-Burke, S., Burke, M. D., Martin, E., Boon, R. T., Fore, C., III, & Kirkendoll, D. (2005). The internal consistency of the school-wide subscales of the Effective Behavioral Support Survey. *Education and Treatment of Children*, *28*, 400–413.

Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., . . . Esparanza, 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. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Kim, J., McIntosh, K., & Hoselton, R. (2014). *Do schools with adequate Tier I SWPBIS implementation have stronger implementation at Tiers II and III?* (PBIS Evaluation Brief). Eugene: OSEP National Technical Assistance Center on Positive Behavioral Interventions and Supports, University of Oregon.

Kincaid, D., Childs, K., & George, H. P. (2005). *School-wide benchmarks of quality* (Unpublished instrument). Tampa: University of South Florida.

Lee, J., & Reeves, T. (2012). Revisiting the impact of NCLB high-stakes school accountability, capacity, and resources: State NAEP 1990–2009 reading and math achievement gaps and trends. *Educational Evaluation and Policy Analysis*, *34*, 209–231. doi:10.3102/0162373711431604

McIntosh, K., & Goodman, S. (2016). *Integrated multi-tiered systems of support: Blending RTI and PBIS*. New York, NY: Guilford Press.

McIntosh, K., Mercer, S. H., Nese, R. N. T., Strickland-Cohen, M. K., & Hoselton, R. (2016). Predictors of sustained implementation of school-wide positive behavioral interventions and supports. *Journal of Positive Behavior Interventions*, *18*, 209–218. doi:10.1177/1098300715599737

McIntosh, K., Mercer, S. H., Nese, R. N. T., Strickland-Cohen, M. K., Kittelman, A., Hoselton, R., . . . Horner, R. H. (2017). *Factors predicting sustained implementation of a universal behavior support framework*. Manuscript submitted for publication.

McIntosh, K., Predy, L. K., Upreti, G., Hume, A. E., Turri, M. G., & Mathews, S. (2014). Perceptions of contextual features related to implementation and sustainability of school-wide positive behavior support. *Journal of Positive Behavior Interventions*, *16*, 31–43. doi:10.1177/1098300712470723

Mercer, S. H., McIntosh, K., & Hoselton, R. (2017). Comparability of fidelity measures for assessing tier 1 school-wide positive behavioral interventions and supports. *Journal of Positive Behavior Interventions*, *19*, 195–204. doi:10.1177/1098300717693384

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide: Statistical analysis with latent variables: User's guide* (7th ed.). Los Angeles, CA: Author.

Muthén, L. K., & Muthén, B. O. (2009, March). *Mplus short courses topic 8: Multilevel modeling with latent variables using Mplus: Longitudinal analysis* (Handout in the Mplus Short Course Video). Retrieved from https://www.statmodel.com/download/Topic%208-%20v17.pdf

National Center for Education Statistics. (2014). *Public elementary/ secondary school universe survey data, 2012-13* [Database]. Retrieved from http://nces.ed.gov/ccd/

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, *78*, 33–84. doi:10.3102/0034654307313793

Pas, E. T., & Bradshaw, C. P. (2012). Examining the association between implementation and outcomes. *The Journal of Behavioral Health Services & Research*, *39*, 417–433.

Pas, E. T., Waasdorp, T. E., & Bradshaw, C. (2015). Examining contextual influences on classroom-based implementation of positive behavior support strategies: Findings from a randomized controlled effectiveness trial. *Prevention Science*, *16*, 1096–1106.

Reinke, W. M., Herman, K. C., & Stormont, M. (2013). Classroom-level positive behavior supports in schools implementing SW-PBIS: Identifying areas for enhancement. *Journal of Positive Behavior Interventions*, *15*, 39–50. doi:10.1177/1098300712459079

Simonsen, B., Eber, L., Black, A. C., Sugai, G., Lewandowski, H., Sims, B., & Myers, D. (2012). Illinois statewide positive behavioral interventions and supports: Evolution and impact on student outcomes across years. *Journal of Positive Behavior Interventions*, *14*, 5–16. doi:10.1177/1098300711412601

Solomon, B. G., Tobin, K. G., & Schutte, G. M. (2015). Examining the reliability and validity of the effective behavior support self-assessment survey. *Education and Treatment of Children*, *38*, 175–191.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, *42*, 893–898. doi:10.1016/j.paid.2006.09.017

Sugai, G., Horner, R. H., & Todd, A. W. (2000). *PBIS Self Assessment Survey 2.0*. Eugene, OR: OSEP Center for Positive Behavioral Supports.

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

Sugai, G., Lewis-Palmer, T., Todd, A. W., & Horner, R. H. (2001). *School-wide evaluation tool* (Version 2.0). Eugene: Educational and Community Supports, University of Oregon.

Sugai, G., Todd, A. W., & Horner, R. H. (2001). *Team implementation checklists* (Version 2.2). Eugene, OR: OSEP Center for Positive Behavioral Supports.

Turri, M. G., Mercer, S. H., McIntosh, K., Nese, R. N., Strickland-Cohen, M. K., & Hoselton, R. (2016). Examining barriers to sustained implementation of school-wide prevention practices. *Assessment for Effective Intervention*, *42*, 6–17. doi:10.1177/1534508416634624