Received: 18 August 2019 | Revised: 11 December 2019 | Accepted: 16 December 2019

DOI: 10.1002/pits.22336

RESEARCH ARTICLE

WILEY

Updating and expanding systematic reviews and meta-analyses on the effects of school-wide positive behavior interventions and supports

Ahhyun Lee | Nicholas A. Gage

School of Special Education, School Psychology, and Early Childhood Studies, College of Education, University of Florida, Gainesville, Florida

Correspondence
Nicholas A. Gage, School of Special Education, School Psychology, & Early Childhood Studies, University of Florida, 1403 Norman Hall, PO Box 117050 Gainesville, FL 32611.
Email: gagenicholas@coe.ufl.edu

Abstract

School-wide positive behavior interventions and supports (SWPBIS) is a multitiered support framework for preventing problem behaviors and increasing prosocial behaviors. There has been an increasing number of experimental group design research studies focused on the impacts of SWPBIS. Therefore, we conducted a systematic review to update and extend prior reviews by including both published and unpublished studies and conducting a robust variance estimation meta-analysis to examine the magnitude of the effect of SWPBIS on all available school-level outcomes. A total of 29 studies were identified across the United States and Europe, including seven randomized controlled trials and 22 quasi-experimental group design studies. Overall, statistically significant reductions in school discipline and increased academic achievement were found. Effect sizes were small to medium. Limitations and implications are discussed.

KEYWORDS

academic achievement, meta-analysis, organizational health, school discipline, school-level outcomes, school-wide positive behavior intervention and support, systematic review

1 | INTRODUCTION

A series of narrative and systematic reviews and meta-analyses examining the effects of school-wide positive behavior interventions and supports (SWPBIS) on school, staff, and student outcomes have been published over the past decade. Recently, there has been a noticeable increase in the quantity and quality of published experimental

group design studies focused on SWPBIS (e.g., Grasley-Boy, Gage, & Lombardo, 2019; Ryoo, Hong, Bart, Shin, & Bradshaw, 2018). Therefore, this study updates those prior reviews and, importantly, addresses key limitations of previous reviews by (a) including unpublished research and (b) conducting a robust variance estimation (RVE) meta-analysis to address the dependence of multiple effect sizes reported in each study.

1.1 | SWPBIS defined

SWPBIS is a multitiered system of support framework designed to prevent problem behavior from occurring and effectively intervening when it does (Sugai & Horner, 2009). Critical features of SWPBIS include school-based teams making data-based decisions, identifying students in need of increased intervention intensity, which is delivered via tiers of prevention and intervention (Horner, Sugai, & Anderson, 2010). Primary prevention, or Tier 1, provides universal supports for all students, including establishing and teaching school-wide behavioral expectations, teaching school routines, increasing the use of evidence-based classroom management, and developing a school-wide recognition system for appropriate behavior (McIntosh & Goodman, 2016). Secondary interventions, or Tier 2, are typically implemented with a small group of students and include strategies such as Check and Connect (Anderson, Christenson, Sinclair, & Lehr, 2004), Check-in/Check-out (Crone, Hawken, & Horner, 2010), and Social Skills Groups (Lane et al., 2003). Lastly, tertiary intervention, or Tier 3, typically includes functional behavior assessment-based behavioral intervention plans to target the most intensive behavioral needs of individual students (Horner, Kincaid, Sugai et al., 2014).

SWPBIS is not a curriculum, but rather a framework promoting a process of the capacity building between professionals along with continuous engagement in professional development (Horner et al., 2010). Sources of professional development are variable across states but are often provided from state or district personnel to ensure fidelity of implementation. Fidelity assessment is important to (a) identify strengths and weaknesses of the implementation process, (b) evaluate the effectiveness of adopted programs, and (c) determine student progress as it takes time to emerge (McIntosh & Goodman, 2016). Currently, there are more than 25,000 schools worldwide implementing SWPBIS with varying degrees of fidelity (www.pbis.org).

1.2 | Prior reviews of SWPBIS

To date, three systematic reviews (i.e., replicable search and inclusion/exclusion procedures) have examined the experimental evidence of the impact SWPBIS has on school, staff, and student outcomes. Solomon, Klein, Hintze, Cressey, and Peller (2012) reviewed single-case designs and identified 20 studies that included at least one individual component of SWPBIS as an independent variable and student behavior as a dependent variable. The authors used a regression-based procedure to calculate single-case design effect sizes and found a small effect (r^2 = .35). However, the authors included pre–post studies, which did not demonstrate experimental control, and the levels of dependent variables varied from student- to classroom- to school-levels (i.e., treating student-level effects equivalent to school-level effects). In addition, the fidelity of implementation data was not reported, and evaluation of study quality was absent.

Gage, Whitford, and Katsiyannis (2018) conducted a systematic review focused exclusively on group-experimental design studies that examined the effect of SWPBIS implementation on disciplinary exclusions (i.e., office discipline referrals [ODRs] and suspensions) at the school level. They identified four studies and reviewed the quality of those studies using the What Works Clearinghouse (WWC) standards and the Council for Exceptional Children (CEC) quality indicator guidelines. The authors found that the overall treatment effect on combined ODRs and suspensions was not statistically significant. However, the authors did find the effect of SWPBIS on school suspensions alone was large and statistically significant (g = −0.86). A limitation of their review was that they focused on exclusionary outcomes and included only a small number of studies due to their strict inclusion criteria.

Most importantly, they only included peer-reviewed and published research studies. Research suggests that exclusion of unpublished research in meta-analyses may bias the estimate of effect (Gage, Cook, & Reichow, 2017).

Similarly, Mitchell, Hatton, and Lewis (2018) conducted a systematic review of group-experimental design studies of SWPBIS, but they focused on SWPBIS training as an independent variable instead of just implementation of SWPBIS. As a result, the authors identified five studies, which they evaluated using WWC and CEC quality standards. Mitchell et al. (2018) found that schools can be successfully trained to implement SWPBIS with fidelity, which also leads to positive distal effects on school-wide outcomes (e.g., school climate and student behaviors). However, studies did not report the fidelity of SWPBIS training, which was a primary focus of their review, and the authors did not report effect sizes. Furthermore, like Gage et al. (2018), Mitchell et al. (2018), and also excluded unpublished research.

1.3 | Purpose

Despite the growing number of schools implementing SWPBIS, current systematic reviews synthesizing the evidence include only a few experimental studies. Furthermore, those reviews (a) did not synthesize all available outcomes and (b) excluded unpublished research (e.g., dissertations). Therefore, we conducted a systematic review to extend previous syntheses focusing on group-experimental studies that examined the implementation of SWPBIS (at least tier 1) on student outcomes at the school-level by addressing these two limitations. Hence, in contrast to the study focused on exclusionary discipline (Gage et al., 2018), the current review includes all possible school, staff, and student-level outcomes, including academic, behavioral, and organizational outcomes. In addition, this review includes dissertations and unpublished studies to reduce potential publication bias on the results (Gage, Leite, Childs, & Kincaid, 2017). The following research questions guided this review:

1. How many group-experimental (i.e., randomized control trials) or quasi-experimental design (QED) studies examined the effect of SWPBIS on school-level outcome?
2. What is the quality of those studies based on WWC and CEC standards?
3. What is the effect of SWPBIS across all included outcomes?

2 | METHODS

We conducted a systematic review to synthesize the effects of SWPBIS on school, staff, and student outcomes by searching electronic databases, hand searching journals, conducting forward and backward searches, and reaching out to authors publishing SWPBIS research. The following steps were used to identify potential studies for inclusion: (a) Abstract review from the electronic database search, (b) full-text review, and (c) full-text coding and data extraction, conducted two times for reliability. Then, a hand search of the *Journal of Positive Behavioral Interventions* and a forward and backward (or ancestral) search were conducted from the collected peer-reviewed journal. Last, we reached out to authors identified in the search about current, unpublished, or submitted research.

2.1 | Search procedures

2.1.1 | Abstract search

Multiple search strategies to identify all studies meeting the inclusion criteria were used. These strategies included searching of electronic databases and checking reference lists from included studies meeting eligibility criteria. The

electronic database search included Education Resources Information Center (ERIC) and Academic Search Premier, Education Index Retrospective, Education Source, PsycINFO, Psychology and Behavioral Sciences Collection in EBSCO host, and ProQuest Dissertations and Theses Global in October, 2019. The following terms were used for in a Boolean/Phase search: ("positive behavioral intervention* and support*" OR "positive behavior* support*" OR "multi-tier" OR "multitier" OR "multi tier") AND ("outcome" OR "academic*" OR "behavior*" OR "achieve*" OR "perform*") AND ("affect*" OR "comparison group" OR "control*" OR "effect*" OR "experiment*" OR "impact*" OR "QED" OR "Quasi-experimental" OR "quasiexperimental" OR "random*" OR "RCT" OR "treatment"). The total of 1,242 abstracts in EBSCO, 518 in ERIC, and 529 in and ProQuest Dissertations and Theses Global were identified through the database search and exported to *Endnote*. After the removal of duplicated records (137), a total 1,711 records were included for review.

The abstracts were then reviewed to screen for inclusion eligibility using the following criteria: (a) The study was conducted in K-12 public schools; (b) the independent variable was implementation of SWPBIS; (c) the primary outcomes could be assessed at school-level; and (d) studies used group experimental or QED with a true comparison group, excluding case studies (i.e., pre–post not comparison group), single-case design, literature reviews, or meta-analyses, and correlational studies (i.e., within-group longitudinal designs). Studies with one-to-one school comparisons (e.g., Caldarella, Shatzer, Gray, Young, & Young, 2011) were excluded because school-level variance could not be calculated. Also, studies with proprietary models (e.g., Safe and Civil Schools) were excluded (Ward & Gersten, 2013) to examine the effect of the SWPBIS framework as described by Sugai and Horner (2009).

2.1.2 | Full-text review

Thirty-four studies met inclusion criteria from the abstract review and were passed on for the full-text. After the full-text review using the same inclusion criteria outlined above, 13 studies remained. Twenty-one studies were removed due to study design (n = 14), intervention criteria (n = 2), and outcome criteria (n = 5). Studies that focused on specific interventions implemented as part of SWPBIS implementation (e.g., Benner, Nelson, Sanders, & Ralston, 2012; Wills, Kamps, Abbott, Bannister, & Kaufman, 2010) or those that focused exclusively on teacher outcomes (e.g., teacher self-efficacy) were removed (e.g., Ross, Romer, & Horner, 2012). Also, studies with control groups that also received SWPBIS training (e.g., Simonsen et al., 2012; Vincent, Randall, Cartledge, Tobin, & Swain-Bradway, 2011) were excluded. Two studies that examined the same outcomes with the same sample but at different periods, one with a 3-year trial (Bradshaw, Koth, Bevans, Ialongo, & Leaf, 2008) and one with a 5-year trial (Bradshaw, Koth, Thornton, & Leaf, 2009). We include only the 5-year trial.

Next, 45 studies were identified for full-text review from ProQuest Dissertations and Theses Global. Eleven studies remained after the removal of 34 studies due to study design (n = 21), outcome criteria (n = 12), and setting (n = 1). Studies that included researcher selected classrooms to represent school-level outcomes (e.g., Eacho, 2013) or examined cohort effects, not school effects (e.g., Smolkowski, 2006) were excluded. Twenty studies were identified from these two sets of electronic searches with the keywords. The hand search of the past 6 years (2013–2019) of the *Journal of Positive Behavioral Interventions* identified two additional studies. Then, forward (n = 2,404) and backward searches (n = 677) were conducted using the same eligibility criteria. Five additional studies were included (three peer-reviewed journals and two dissertations). Last, we contacted authors of included studies to identify additional, unpublished studies, which resulted in five additional studies (four peer-reviewed and one dissertation). Overall, we identified 20 peer-reviewed studies and 12 dissertations (k = 32).

2.1.3 | Full-text coding and data extraction

We extracted study characteristics in *Excel* using the following criteria: (a) Study design, (b) setting, (c) number of schools, (d) number of students, (e) school/student characteristic, (f) the number of years implementing SWPBIS, (g)

implementation levels of SWPBIS, (h) implementation fidelity, (i) areas of outcomes, and (j) outcome measures. Then, the quality of each study was assessed using both the WWC Group Design Standards 4.0 (2019) and the CEC quality indicators (CEC, 2014; Cook et al., 2015). For the WWC Standards, studies were categorized into one of three ratings: (a) *Meets WWC Group Design Standards Without Reservations*, (b) *Meets WWC Group Design Standards with Reservations*, or (c) *Does not Meet WWC Group Design Standards*. Studies had to meet three criteria to meet without reservations: (a) A randomized controlled trial (RCT) with, (b) no risk of bias due to individuals entering clusters (i.e., joiners), and (c) no risk of bias due to nonresponse (i.e., attrition). RCTs with high attrition or QED studies could meet with reservations if baseline equivalence was established (see WWC, 2019 for a complete description of standards).

The CEC quality indicators (2014) include eight domains: (a) Context and setting, (b) participants, (c) intervention agent, (d) description of practice, (e) implementation fidelity, (f) internal validity, (g) outcome measures, and (h) data analysis. The total of 24 items for group design studies was considered and coded as being (a) present or (b) not present. We then summed the number of present items and divided by a total of 24 items to calculate the percentage of indicators met according to the CEC quality indicator.

2.2 | Inter-rater agreement

Two trained raters coded abstracts and studies, extracted data, and conducted quality assessments. The first author conducted 100% of all data collection phases. The second author conducted an inter-rater agreement. Any disagreements were resolved using a consensus meeting, where item disagreements were discussed together and a final decision for the final value was determined. Inter-rater agreement was conducted for 10% of all abstracts were double coded, with 100% agreement for passing on to the full-text review. Fifty-percent of full-text review articles were double coded, with 100% agreement. Fifty-percent of the final included studies were double coded, with 98% agreement across all items. A consensus meeting was conducted to resolve discrepancies. Finally, 30% of the included studies' quality was double coded. One-hundred percent agreement was found for both study quality instruments.

2.3 | Meta-analytic procedures

First, we extracted means, standard deviations (*SD*), proportions (e.g., percentage of students in a school with one or more ODR), and sample sizes for all included outcomes. If descriptive statistics were not available, we extracted *t* values for *t* tests (suspensions in Algozzine, Wang, White et al., 2012), *F* values from analysis of variance models (Guardino, 2013), or recorded the author reported effect sizes (Bradshaw, Mitchell, & Leaf, 2010; Bradshaw, Waasdorp, & Leaf, 2012; Gage, Lee et al., 2018; Gage, Lee, Grasley-Boy, & Peshak George, 2018; Gage & Stevens, 2018; Grasley-Boy et al., 2019). We then calculated standardized mean difference effect sizes (*g*) and the variance of *g* for all outcomes following formulas described in Lipsey and Wilson (2001). The unit of interest for this study was the school-level effect of SWPBIS. Therefore, for all studies, we used the number of schools to calculate the variance of *g*, even when the study reported student-level outcomes. We took this approach to ensure that the weights were accurately applied in the subsequent meta-analysis. For example, Bradshaw et al. (2012) reported behavior outcomes from the Teacher Observation of Classroom Adaptation (TOCA), a teacher completed rating scale, for 11,738 students in 37 schools. If the variance was based on the student-level sample, the weight for the study would be much larger than the other studies that report school-level data, thereby biasing the results toward the reported effect sizes in that study only because student-level data were reported.

Next, we estimated meta-analytic models for behavior, academic, and organizational domains. Most studies reported more than one outcome for the behavior and academic domains (e.g., ODR, in-school suspension [ISS], and

out-of-school suspension [OSS]). Therefore, we estimated RVE meta-analytic models for the behavior and academic outcomes. RVE is a meta-analytic approach used to analyze statistically dependent effect sizes (Tanner-Smith, Tipton, & Polanin, 2016). Multiple effect sizes from the same study and sample are statistically dependent and violate the assumption of independence of traditional meta-analysis models (Borenstein, Hedges, Higgins, & Rothstein, 2009). Only one outcome was reported from each study for the organizational domain, therefore, a traditional random-effects model was estimated. We made an a priori decision to estimate random-effects for all models because we believe that (a) there may be additional studies that we were unable to collect and study-level characteristics may account for between-study variance. For the behavior and academic outcomes, we conducted a series of subgroup analyses to evaluate moderating effects of publication type (dissertation or published study), study design (RCT or QED), and outcomes type (disciplinary exclusions, suspensions, reading, and math). There were not enough effect sizes in the organizational domain to conduct subgroup analyses. All meta-analyses and moderator analyses were conducted in R (R Core Team, 2014). The RVE models were estimated in the *robumeta* package (Fisher, Tipton, & Zhipeng, 2017), while the random-effects model for the organizational domain was estimated in the *metafor* package (Viechtbauer, 2010).

3 | RESULTS

Overall, 32 studies using an RCT or QED were identified that evaluated the effect of SWPBIS on school-level outcomes. Six of the studies used an RCT design (Algozzine et al., 2012; Bradshaw et al., 2010, 2012; Horner, Sugai, Smolkowski, et al., 2009; Sørlie & Ogden, 2015; Waasdorp, Bradshaw & Leaf, 2012), while 26 studies used a QED design with a comparison group that did not receive SWPBIS training. Characteristics of the collected studies are presented in Table 1. Outcomes of the studies were coded by domain. It is worth noting that Horner et al. (2009) reported all three categories, but the behavioral outcomes were not included since there was no control group for ODRs. The authors did not report ODR results for year one, only after delayed treatment was delivered.

3.1 | Study characteristics

A total of 8,781 schools were included across ten different locations in the United States and in Europe. Studies were conducted in the Pacific Northwest, Midwest, Mid-Atlantic region, Southeastern, and Southwestern United States. Specific states included Texas, California, Connecticut, Florida, Georgia, Hawaii, Illinois, Kentucky, Maryland, Minnesota, Missouri, North Carolina, Oregon, Pennsylvania, and South Carolina, while one study was conducted in Norway (Sørlie & Ogden, 2015).

Sixteen studies were conducted in elementary schools, three in middle school, two in high school, and 11 included mixed levels of schools with the majority at the elementary level. The average student enrollment was 698 students per school. Student characteristics were varied. The percentage of White students ranged from 8% to 86%, while the percentage of African American students ranged from 4% to 59%, while the percentage of students receiving free or reduced lunch (FRL) ranged between 9% and 69% across the samples.

In regard to the implementation levels of SWPBIS, all but three of the studies focused on the implementation of Tier 1. Two studies evaluated the implementation of all three tiers (Algozzine et al., 2012; Gage, Grasely-Boy et al., 2019) while one study evaluated the universal and tertiary tiers (Nelson, Martella, & Marchand-Martella, 2002). Across the studies, the average years of implementation were 3.4 years. In terms of the tools used to evaluate the fidelity of implementation, most used the School-wide Evaluation Tool (SET; Horner, Todd, Lewis-Palmer, et al., 2004) or the Benchmark of Quality (BoQ; Cohen, Kincaid, & Childs, 2007). Seven studies did not report the fidelity tool used. Overall, reported fidelity for the treatment groups were at or above 70% on the BoQ and 80% on the SET. Ten studies did not report fidelity scores.

TABLE 1 Characteristics of the included studies

Study (n = 32)	Study design	Setting	No. of schools	No. of students	Student characteristics	No. of years SWPBIS	Levels of SWPBIS	Fidelity
Algozzine et al. (2012)	RCT	Elementary schools in a large school district in North Carolina	6 Schools (4 treatment)	Average enrollment = ~500 per school	58.6% Black; 64.0% FRL; 11.4% SPED	1 Year	All 3 tiers	SET = 93.8%
Arnold (2012)	QED	Middle schools in one northwest county in Georgia	3 Schools (2 treatment)	Average enrollment = ~900 per school	9.66% Black; 58.6% FRL; 14.6% SPED	~4.5 Years	Tier 1	SET = 91%
Bradshaw et al. (2009, 2010)	RCT	Elementary schools in five different districts (48% suburban, 41% urban, and 11% rural) in Maryland	37 Schools (21 treatment)	Average enrollment = ~488 per school	60.7% White; 39.6% FRL; 14.1% SPED	4 Years	Tier 1	SET = ~95%
Bradshaw et al. (2012); Waasdorp et al. (2012)	RCT	Elementary schools in five different districts (48% suburban, 41% urban, and 11% rural) in Maryland	37 Schools (21 treatment)	Average enrollment = ~486 per school	45.1% Black; 46.1% White; 12.9% SPED; 49.4% FRL	4 Years	Tier 1	SET ≥ 80%
Darwin (2016)	QED	Elementary and middle schools in two suburban school districts in the Southwestern	8 Schools (6 treatment)	Unclear	Treatment (43.6% White; 8.9% Black; 47.0% FLR; 8.0% SPED) control (26.8% White; 14.75% Black; 25.05% FLR; 9% SPED)	Unclear	Tier 1	BoQ; SET ≥ 88%; EBS
Dirsmith (2013)	QED	Elementary schools in Pennsylvania	26 Schools (13 treatment)	Average enrollment = ~513 per school	30.0% Minority; 45% ED; ~13.6% SPED	4 Years	Tier 1	SET = 86.3%; BoQ = 89.3%; TIC = 86.3%

(Continues)

TABLE 1 (Continued)

Study ($n = 32$)	Study design	Setting	No. of schools	No. of students	Student characteristics	No. of years SWPBIS	Levels of SWPBIS	Fidelity
Farrington (2018)	QED	Grade 3 students in one South Florida school district	277 Schools (30 treatment)	Unclear	76.6% FRL	Unclear	Unclear	Unclear
Flannery, Fenning, Kato, and McIntosh (2014)	QED	High schools in the Pacific Northwest and Midwest	12 Schools (8 treatment)	Average enrollment = ~1,770 per school	30.0% non-White; 27.0% FRL	3 Years	Tier 1	SET collected, but fidelity not reported
Gage, Grasley-Boy et al. (2018)	QED	Public schools in Florida (71% elementary, 19% middle, and 9% high)	1,186 Schools (593 treatment)	Average enrollment = ~828 per school	48.9% White; 20.7% Black; 23.7% Hispanic; 60.7% FRL	4.5 Years	Tier 1	BoQ ≥ 70%
Gage, Grasley-Boy et al. (2019)	QED	Public elementary schools in California (78% elementary, 10.5% middle, and 11% high)	196 Schools (98 treatment)	Average enrollment = ~749 per school	15.2% White; 3.8% Black; 70.3% Hispanic; 75.2% FRL; 10.3% SPED	Unclear	All 3 tiers	TFI ≥ 70%
Gage, Lee et al. (2018)	QED	Public elementary and intermediate schools in Georgia	Total 238 schools (119 treatment)	Average enrollment = ~529 per school	44.1% White; 32.6% Black; 16.2% Latino; 68.9% ED; 11.2% SPED	Unclear	Tier 1	BoQ ≥ 70%
Gage et al. (2017)	QED	All public elementary schools in the state of Florida	~2,033 Schools, including an average of 235 schools implementing SWPBIS with fidelity each year	Average enrollment = ~650 per school	50.0% White; 23.0% Black; 23.0% Hispanic; 54.0% FRL	Average 5.7 years	Tier 1	BoQ = 83.6%

(Continues)

TABLE 1 (Continued)

Study ($n = 32$)	Study design	Setting	No. of schools	No. of students	Student characteristics	No. of years SWPBIS	Levels of SWPBIS	Fidelity
Gage, Rose et al. (2019)	QED	Public elementary and intermediate schools in Georgia	236 Schools (118 treatment)	Average enrollment = ~649 per school	44.7% White; 33.0% Black; 15.3% Hispanic; 69.2% ED; 11.2% SPED	Unclear	Tier 1	BoQ ≥ 70%
Gage and Stevens (2018)	QED	Public schools in South Carolina (65% elementary, 26% middle, and 7% high)	270 Schools (135 treatment)	Average enrollment = ~710 per school	44% Black; 8% White; 59% FRL	NA	Tier 1	SET ≥ 80%
Gage, Sugai, Lewis, and Brzozowy (2015)	QED	Public schools in Connecticut (64% elementary, 21% middle, 14% high, 7%)	300 Schools (150 treatment)	Average enrollment = ~558 per school	59.0% White; 18.0% Black; 19.0% Hispanic; 41.0% FRL	Average 3.11 years	Tier 1	SET = 94%
Grasley-Boy et al. (2019)	QED	Public schools in California (78% elementary, 11% middle, and 11% high)	Total 1,088 (544 treatment)	Average enrollment = ~597 per school	17.9% White; 7.3% Black; 57.4% Hispanic; 10.3% SWD; 69.2% FRL	Unclear	Tier 1	TFI ≥ 70%
Guardino (2013)	QED	Middle schools in Oregon	51 Schools (34 treatment)	Average enrollment = ~610 per school	72.0% White; 14.0% SWD; 46.0% FRL	3 Years	Tier 1	SET; BoQ; TIC; EBS-SAS; but scores were unclear
Hirschi (2015)	QED	Middle schools in Missouri	4 Schools (2 treatment)	Unclear	40.0% ≤ FRL ≤ 60.0%	≥2 Years	Tier 1	Unclear
Homen (2014)	QED	Elementary schools in a suburban school district in the South Bay area of Northern California	13 Schools (10 treatment)	Average enrollment = ~166 per school	52.9% Hispanic; 26.3% White; 12.9% Asian; 4.2% Black; 7.8% SPED	1–2 Years	Unclear	Unclear

(Continues)

TABLE 1 (Continued)

Study (n = 32)	Study design	Setting	No. of schools	No. of students	Student characteristics	No. of years SWPBIS	Levels of SWPBIS	Fidelity
Horner et al. (2009)	RCT	Elementary schools in Hawaii and Illinois	63 Schools (33 treatment)	Average enrollment = ~471 per school	61.0% non-White; 51.0% FRL; 9.0% IEP	3 Years	Tier 1	SET ≥ 80%
Houchens et al. (2017)	QED	Public schools in Kentucky (62% elementary, 21% middle, and 17% high)	295 Schools (151 treatment)	Average enrollment = ~459 per school	85.0% White; 57.0% FRL	Average 3 years	Tier 1	BoQ (low: ≤ 70%; medium: 71–89%; high: ≥ 90%)
Mozley (2016)	QED	Public schools in ~90% of metropolitan in a Southeastern state (64% elementary, 24% middle, and 12% high)	256 Schools (128 treatment)	Average enrollment = ~817 per school	61.0% FRL	≥3 Years	Tier 1	BoQ ≥ 70%
Nelson et al. (2002)	QED	Elementary schools in Pacific Northwest	35 Schools (7 treatment)	Average enrollment = ~492 per school	86.0% White; 55.0% FRL	2 Years	Tier 1 and 3	Unclear
Pas et al. (2019)	QED	Elementary, middle, and high Schools in Maryland	1,316 Schools (859 treatment)	Average enrollment elementary = ~477 per school; average enrollment secondary = ~1,119 per school	48.9% White; 38.3% Black; 37.3% FRL; 11.9% SPED	6 Years	Tier 1	SET ≥ 80%; IPI ≥ 80%
Porter (2012)	QED	Elementary schools in an urban/suburban district in the Mid-Atlantic region	114 Schools (57 treatment)	Average enrollment = ~641 per school	43.6% White; 10.1% Black; 24.8% FRL; 49.5% emotional disturbance	Unclear	Tier 1	SET ≥ 80%; BoQ (full = 86.9%; partial = 58.1%)
Rhodes-Monette (2014)	QED	High schools in a suburban school district in Texas	10 Schools (8 treatment)	Average enrollment = ~2,968 per school	41.0% White; 16.2% Black; 31.6% ED; 7.5% SPED	Average 3 years	Tier 1	Unclear

(Continues)

TABLE 1 (Continued)

Study (n = 32)	Study design	Setting	No. of schools	No. of students	Student characteristics	No. of years SWPBIS	Levels of SWPBIS	Fidelity
Ryoo et al. (2018)	QED	Elementary and middle school in Minnesota (32 elementary and 34 middle schools)	66 Schools (33 treatment)	Unclear	Unclear	Unclear	Tier 1	SET was used, but scores were unclear
Schultz (2007)	QED	Public schools in Texas (60% elementary, 22% junior high/middle, and 18% high schools)	103 Schools (56 treatment)	Unclear	Unclear	3 Years	Tier 1	Unclear
Sørlie and Ogden (2015)	RCT	Primary schools (Grade 1–7) in Norway	48 Schools (28 treatment)	Average enrollment = ~282 per school	Unclear	4 Years	Tier 1	EBS-SAS ≥ 80% (75% of schools in treatment)
Taylor (2017)	QED	Elementary schools in Kentucky (Grade 5 only)	454 Schools (112 treatment)	Average enrollment = ~74 per school	60.5% low-SES	3 Years	Tier 1	BoQ ≥ 70%
Waasdorp et al. (2012)	RCT	Elementary Schools in Maryland	37 Schools (21 treatment)	Median enrollment = 359 per school	46.1% White 45.1% Black 49.5% FRL 12.9% SPED	4 Years	Tier 1	SET was used, but scores were unclear

Abbreviations: BoQ, Benchmark of Quality; EBS, The Effective Behavior Support Survey; EBS-SAS, Effective Behavior Support Self-Assessment Survey (Sugai, Horner, & Todd, 2003); ED, economically disadvantaged; FRL, free and reduced lunch; IEP, individualized education program; IPI, Implementation Phases Inventory; RCT, randomized controlled trial; QED, quasi-experimental design; SES, social economic status ; SET, School-wide Evaluation Tool; SPED, students receiving special education services; TIC, Team Implementation Checklist (Sugai et al., 2002); TFI, Tiered Fidelity Inventory (TFI; Algozzine et al., 2014).

Twenty peer-reviewed published studies and 12 dissertations were included in this review. When comparing results by publication type (i.e., peer-reviewed journal and dissertation), the average number of schools included in peer-reviewed journals was average 394.7 schools (total 7,499 schools), whereas the average in dissertations was 109.9 schools (total 1,319 schools). Regarding study design, one-third of the studies in peer-reviewed journals were RCT, while none of the dissertations used an RCT.

3.2 | Quality assessment based on WWC and CEC standards

Nineteen studies met WWC standards. Overall, 51.7% of the peer-reviewed journal articles met WWC standards with or without reservations ($k = 17$), while 16.6% of the dissertations met the WWC with or without standards ($k = 2$). Seven studies met WWC standards without reservation, 12 studies met WWC standards with reservation by establishing equivalence for both groups, while the remaining 12 studies did not meet the WWC standards. For studies that did not have prior year behavioral outcomes, primarily due to limited information about how long schools had been implementing, equivalence was established for all available school characteristics, including ethnicity and socioeconomic status.

Five of the seven studies that met WWC standards without reservation also met all of the CEC quality indicators, while the other two studies meeting without reservations (Sørlie & Ogden, 2015; Waasdorp, Bradshaw, & Leaf, 2012) met ~80% of the quality indicators. Across all studies, the percentage of CEC quality indicators met ranged from 38% to 100%, with an average of 75% met. The average percentage of indicators met in peer-reviewed journals was 83%, while that in dissertations was 63% (contact author for complete tables).

3.3 | Study reported effects of SWPBIS on school-level outcomes

Twenty-two of the 32 studies reported outcomes in the behavior domain, 13 studies reported outcomes in the academic domain, and four studies reported outcomes in the organizational domain (see Table 2). Sixteen studies reporting outcomes in the behavior domain included suspensions (ISS and OSS) and nine studies included ODRs. Other behaviors outcomes included corporal punishment, expulsion, referral to law enforcement, school-related arrest (e.g., Gage, Grasley-Boy, Peshak George, Childs, & Kincaid, 2018; Grasley-Boy et al., 2019), bullying and peer rejection (Gage, Rose et al. 2019; Waasdorp et al., 2012), and emotional disturbance eligibility (Porter, 2012). Most studies reported proportions of students with a behavior outcome (e.g., OSS), while two studies used the TOCA-checklist (TOCA-C; Koth, Bradshaw, & Leaf, 2009). Results from the peer-reviewed articles indicate that there were statistically significant decreases in behavior domain outcomes, primarily for ODRs and suspension, but also for other more severe discipline resulting from problem behaviors (e.g., referral to law enforcement and arrest). Three studies in peer-reviewed journals reported null results for behavior (Gage, Rose, et al., 2019; Gage & Stevens, 2018; Ryoo et al., 2018). The results from the dissertations were mixed (i.e., significant and null) for both ODRs and suspensions, but significant results were found for emotional disturbance eligibility, specifically, schools implementing SWPBIS reported fewer students with emotional disturbance (Porter, 2012).

For outcomes in the academic domain, most studies included reading, mathematics, or aggregated both. Four studies reported additional subjects, including science, social studies, and writing. Eleven studies used state summative tests from Connecticut, Florida, Hawaii and Illinois, Kentucky, Maryland, Minnesota, North Carolina, Pennsylvania, and Texas. All of the studies reporting state summative test results reported the percentage of students at or above proficient in the achievement domain (e.g., reading). Overall, the results from peer-reviewed journals had mixed results (i.e., significant and null); all results from dissertations were null.

Last, four studies examined organizational outcomes. Three studies examined student perceptions of school safety using the Student Safety Survey (Spokane Public District, 1997). Nelson et al. (2002) and Horner et al. (2009)

TABLE 2 Outcomes of the included studies

Study (n = 32)	Academic area	Outcome measures	Behavior area	Outcome measures	Organizational area	Outcome measures
Algozzine et al. (2012)	Reading	DIBELS, percentage at or above proficient on State Reading Standard	Exclusion	ODR and suspension		
Arnold (2012)			Exclusion	ODR, ISS, OSS, expulsion, and arrest		
Bradshaw et al. (2009)					Teacher report of organizational health	OHI
Bradshaw et al. (2010)	Reading and math	Percentage at or above proficient on Maryland School Assessment	Exclusion	ODR and suspension		
Bradshaw et al. (2012)			Disruptive behavior, concentration problem, prosocial behavior, emotion regulation, and exclusions	TOCA-C, ODR, and suspension		
Darwin (2016)	Reading, writing, and math	STAAR				
Dirsmith (2013)	Reading and math	Percentage at or above proficient on PSSA and PVAAS				
Farrington (2018)	Reading	Percentage at or above proficient on Florida Standards Assessment				
Flannery et al. (2014)			Exclusions	ODR		
Gage et al. (2017)	Reading and math	Percentage at or above proficient on FCAT				
Gage et al. (2015)	Reading, math, and writing	Percentage at or above proficient on CMT for Grade 3–8 and the CAPT for Grade 10				
Gage, Grasley-boy et al. (2018)			Exclusions and corporal punishment	Corporal punishment, ISS, OSS, expulsion, referral to law enforcement, and school-related arrest		

(Continues)

TABLE 2 (Continued)

Study (n = 32)	Academic area	Outcome measures	Behavior area	Outcome measures	Organizational area	Outcome measures
Gage, Grasley-Boy et al. (2019)			Exclusions	Suspensions and expulsions		
Gage, Lee et al. (2018)			Exclusions	ISS, OSS, and disciplinary incident		
Gage, Rose, and Kramer (2019)			Bullying	School climate survey		
Gage and Stevens (2018)			Exclusions and corporal punishment	ISS, OSS, and corporal punishment		
Grasley-Boy et al. (2019)			Exclusions	ISS, OSS, expulsions, referrals to alternative schools for disciplinary reason, referrals to law enforcement, and school-related arrest		
Guardino (2013)			Exclusions	ISS, OSS, and expulsion		
Hirschi (2015)			Problem behaviors	Fighting, bullying, verbal aggression, etc.		
Homen (2014)					Perceived school safety	SSS
Horner et al. (2009)	Reading	Percentage proficient or above on State Reading Standards			Perceived school safety	SSS
Houchens et al. (2017)	All subjects	Schools' summative performance scores				
Mozley (2016)			Exclusions	Suspension and expulsions		
Nelson et al. (2002)	Language, arts, science, social studies, and math	Average scores of CTBS; CTB and WASL	Exclusions	Suspension and ODR	Perceived school safety	SSS
Pas et al. (2019)	Reading and math	Percentage of students proficient or above on Maryland School Assessments	Exclusions and truancy	Suspensions and truancy		

(Continues)

TABLE 2 (Continued)

Study (n = 32)	Academic area	Outcome measures	Behavior area	Outcome measures	Organizational area	Outcome measures
Porter (2012)			Emotional disturbance eligibility	# Of students identified		
Rhodes-Monette (2014)	Reading	Percentage of students proficient or above on TAKS	Problem behavior and dropout	Discipline infraction, attendance, dropout, and graduation		
Ryoo et al. (2018)	Reading and math	Average scores of MCA-II for Grade 3–8	Exclusions	ISS and OSS		
Schultz (2007)			Exclusions	ISS, OSS, and disciplinary alternative education placement		
Sørlie and Ogden (2015)			Problem behavior	British Scale		
Taylor (2017)	Math	Percentage proficient or above on KPREP				
Waasdorp et al. (2012)			Bullying and peer rejection	TOCA-C		

Abbreviations: British scales (Grey & Sime, 1989); CAPT, Connecticut Academic Performance Test; CMT, Connecticut Mastery Test; CTBS, Comprehensive Test of Basic Skills (CTB/McGraw Hill, 1992); DIBELS, Dynamic Indicators of Basic Early Literacy Skills (Good & Kaminski, 2002); FCAT, Florida Comprehensive Assessment Test; ISS, in-school suspension; KPREP, Kentucky Performance Rating for Educational Progress; MCA-II, Minnesota Comprehensive Assessment -Series II; ODR, office discipline referral; OHI, Organizational Health Inventory for Elementary Schools (Hoy & Feldman, 1996); OSS, out-of-school suspension; PPVAAS, Pennsylvania Value-Added Assessment System; PSSA, Pennsylvania System of School Assessment; Relative risk ratios, African American risk ratio by the white risk ratio; Risk ratios of exclusionary discipline decisions, the number of African American students suspended or expelled divided by the total number of African American students enrolled in schools; SSS, Student Safety Survey (Spokane Public School District, 1997); STAAR, State of Texas Assessment of Academic Readiness; TAKS, The Texas Assessment of Knowledge and Skills; TOCA-C, the teacher observation of classroom adaptation–checklist (Koth et al., 2009); WASL, Washington Assessment of Learning Outcomes (Office of the Superintendent of Instruction, 1996).

found no statistically significant differences between the treatment and comparison groups, while Homen (2014) found a significant increase in ratings on the personal safety factor. Meanwhile, Bradshaw et al. (2009) examined organizational health, resource influence, staff affiliation, academic emphasis, and collegial leadership using the Organizational Health Inventory for Elementary Schools (Hoy & Feldman, 1996). The authors found statistically significant improvements in all but collegial leadership compared with control schools.

3.4 | Meta-analysis

We used RVE meta-analysis to estimate the effect of SWPBIS on behavior and academic achievement and a random-effects model for organizational outcomes. We excluded outcomes that rarely occurred (e.g., expulsions and referral to law enforcement) to ensure accurate comparison between behavioral outcomes. We also excluded the proportion of students identified with EBD (Porter, 2012) because the outcome was not comparable with other behavioral outcomes and student self-report of bullying (Gage, Rose, & Kramer, 2019) because it is conceptually different than behavior incidents. Last, we could not include the outcomes reported in Waasdorp et al. (2012) because no descriptive statistics or effect sizes were reported and the models only evaluated slope values across time, not treatment effects compared with the other studies. We also excluded Pas et al. (2019) because only significant results were reported by year; overall effects were not reported and could not be calculated.

The results for all RVE models, including moderator analyses, are presented in Table 3. Seventeen studies reported 45 outcomes in the behavior domain. The I^2 value suggests that there was significant heterogeneity for the effect sizes, or that 77% of the variance may be explained by moderators. The first model estimated the overall treatment effect for SWPBIS on behavioral outcomes. Results indicate that schools implementing SWPBIS have a statistically significant impact on behavioral outcomes, reducing problem behavior, as measured by disciplinary exclusions and teacher ratings, by 0.26 *SD* units. We then examined a series of moderators to determine if there were differences of magnitude by moderators. None of the moderators were statistically significant, suggesting that there were no differences between dissertations and published studies, RCTs and QEDs, or by outcome type.

Next, we examined the effect of SWPBIS on academic domain outcomes. Thirteen studies contributed 31 effect sizes for the academic domain. The I^2 value suggests that there was no heterogeneity for the 31 effect sizes, or that

TABLE 3 Meta-analysis results

Domain	Moderator	*K*	# of Effect sizes	I^2 (%)	*b*	95% CI	*df*
Behavior							
	Intercept	18	46	76.90	−0.257***	[−0.413, −0.100]	10.8
	Publication type				0.253	[−0.143, 0.650]	5.02
	RCT				0.134	[−0.160, 0.429]	3.61
	Exclusions				−0.149	[−0.551, 0.253]	2.06
	Suspensions				0.005	[−0.184, 0.194]	8.50
Academics							
	Intercept	13	31	0.00	0.113**	[0.020, 0.205]	6.54
	Publication type				−0.201	[−0.469, 0.067]	2.15
	RCT				0.179	[−0.793, 1.152]	1.85
	Reading				0.123	[−0.090, 0.336]	7.5
	Math				−0.041	[−0.239, 0.157]	6.71
Organizational							
	Intercept	4	4	0.00	0.374	[0.021, 0.728]	3

Note: The intercept value is the overall effect size. Models with degree of freedom <4 are untrustworthy. The behavior and academic models were estimated using robust variance estimation meta-analysis. All models were run separately.
Abbreviations: CI, confidence interval; *df*, degree of freedom; RCT, randomized controlled trial.

all of the variance between effect sizes was due to sampling error. The first model reports the overall treatment effect for SWPBIS, which was statistically significant and suggests that schools implementing SWPBIS have 0.11 *SD* units higher achievement. Similar to the behavior domain, none of the moderators were statistically significant. Last, we estimated a random-effects model for the four organizational domain outcomes. Similar to the academic domain model, the I^2 value was 0.00. Although only based on four effect sizes, there was a statistically significant difference between treatment and comparison schools, with schools implementing SWPBIS reporting 0.37 *SD* units higher on the organizational measures.

4 | DISCUSSION

This systematic review and meta-analysis were designed to address two noteworthy limitations of prior reviews of experimental SWPBIS research, namely, inclusion of unpublished research and modeling dependent effect sizes using RVE. We identified 32 RCT or QED studies that included over 8,700 schools and reported outcomes for behavior, academic, and organization outcomes. When reviewing the individual studies, positive results were present for some, but not all, outcomes in the behavior and academics domains. However, after aggregating the effect sizes and applying RVE meta-analysis, we found a statistically significant aggregate effect size for outcomes in both the behavior and academic domains. Further, although the individual studies reported mixed results for the organizational outcomes (i.e., positive and indeterminant results), when aggregated, we found a statistically significant and positive effect. Based on the meta-analytic results from our systematic review, we found that SWPBIS has a statistically significant and meaningful effect on behavior, academics, and organizational health.

When using Cohen's "rule of thumb" for interpreting effect sizes, all of the results would be considered small. However, there has been a call for alternative interpretation frameworks for effect sizes in educational intervention research (Lipsey et al., 2012). Some suggest that effect sizes of 0.20 or 0.25 *SD* units be considered "educationally significant" (Bloom, Hill, Black, & Lipsey, 2008; Hedges & Hedberg, 2007; Lipsey et al., 2012). Kraft (2018) proposes, based on a review of 481 effect sizes from 242 RCTs, that, in education, effect sizes <0.05 be considered small, 0.05 to <0.20 be considered medium, and 0.20 or greater be considered large. Contextualized within this educational effect size framework, the results would suggest that SWPBI has a large effect on behavior and organizational domain outcomes and a medium effect on academic achievement.

This study included all of the studies in the two prior systematic reviews of experimental group-design SWPBIS evidence, but unlike those reviews, found statistically significant effect sizes for outcomes in all three domains. Gage, Lee et al. (2018) found a statistically significant and large effect for suspensions, but the effect size was based on only two studies, while there was no effect for ODR. We found a statistically significant effect size for all behavior outcomes and that there was no difference in the overall effect size for suspensions, ODR, or other behavioral outcomes. As noted, the Mitchell et al. (2018) did not calculate effect sizes, but found generally positive results, which is aligned with our findings here. No prior systematic review or meta-analysis reviewed the effect of SWPBIS on academic achievement. Overall, we believe that the increase in sample size, both the number of studies and the number of schools, increased the accuracy of the population estimate, which we found to be statistically significant.

Although the results indicate that SWPBIS has significant effects across the three domains, it's worth noting that not all studies reported fidelity of implementation data for schools receiving SWPBIS. Fidelity of implementation provides data on how well a school implemented SWPBIS practices, particularly universal Tier 1 practices (McIntosh & Goodman, 2016). Without that data, we cannot say with certainty whether or not the treatment effects may vary by how well schools implement SWPBIS.

Of the studies that met WWC standards 4.0, 14 examined student behavioral outcomes, while seven studies examined student academic achievements, and three investigated organizational outcomes. Most of the studies that met WWC standards and examined behavioral outcomes found positive results, mostly in ODRs and suspensions but also for other more severe behaviors (e.g., referral to law enforcement and arrest), and bullying

and peer rejection. The results for academic achievement and organizational outcomes were mixed, with some finding significant results and other finding null results and those for organizational outcomes. With regard to the CEC quality indicators, only five studies met 100% of indicators. A number of studies did not adequately describe students with disabilities in the sample or the researchers did not directly manipulate the independent variable (e.g., post hoc QEDs).

Although the current review found promising effects of SWPBIS across all three outcomes, only a few reported outcomes for specific population groups (e.g., minority students) or outcomes at the student level. Four studies (Gage, Grasley-Boy, et al., 2018; Grasley-Boy et al., 2019; Mozley, 2016; Porter, 2012) addressed differential effects of SWPBIS on particular groups of students, including students with disabilities (Gage, Grasley-Boy et al., 2018), students identified for emotional disturbance (Porter, 2012), and African American students (Gage, Grasley-Boy et al., 2018; Grasley-Boy et al., 2019; Mozley, 2016). In particular, Gage, Grasley-Boy et al. (2018) found significantly fewer OSS for students with disabilities and African American students in schools implementing SWPBIS compared with students in comparison schools. Porter (2012) found significantly fewer students were identified as having an emotional disturbance. In contrast, Mozley (2016) found no statistical differences for African American students' rates of suspensions and expulsion. Thus, more research is needed to investigate differential effects of SWPBIS for particular groups of students. Understanding differential effects is important to evidence that SWPBIS not only supports the needs of all students but also groups of students considered vulnerable (e.g., students with disabilities). Thus, future studies should consider disaggregating results by gender, ethnicity, and student disability status.

4.1 | Limitation

Although the current review directly addresses noted limitations of previous reviews, several limitations still remain. First, although the search process was rigorous and replicable, there could be missing studies. A great effort was made to identify all potential studies by using several search procedures, including forward and backward searches and hand searches. Second, Algozzine et al. (2012) included both academic and behavioral interventions in their model, therefore, the academic intervention may have also affected students' academic achievement. Third, there were studies that used the same samples, but different outcomes (Bradshaw et al., 2012; Waasdorp et al., 2012), which may increase bias. The RVE modeling was used to directly address the nesting of effect sizes, including repeated samples across studies. Lastly, most studies in this review were focused on the universal implementation of SWPBIS. Future research should consider examining the effects of SWPBIS not only at Tier 1, but also Tiers 2 and 3.

4.2 | Conclusion

The current review synthesizes the growing evidence of SWPBIS effects on behavioral, academic, and organizational outcomes and extends prior reviews (Gage et al., 2018; Mitchell et al., 2018; Solomon et al., 2012) by including unpublished studies and employing a RVE meta-analysis. Overall, we found statistically and educationally significant effects of SWPBIS on all three school-level domains. These results provide further support for the positive impacts of SWPBIS can have on schools and student outcomes.

ORCID

Nicholas A. Gage http://orcid.org/0000-0002-1800-1760

REFERENCES

Studies included in the literature review are marked with an asterisk

Algozzine, B., Barret, S., Eber, L., George, H., Horner, R., Lewis, T., … Sugai, G. (2014). *School-wide PBIS Tiered Fidelity Inventory. OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports*. Retrieved from https://assets-global.website-files.com/5d3725188825e071f1670246/5daf1f12bc3bf471b96cbe2b_SWPBIS%20Tiered%20Fidelity%20Inventory%20(TFI).pdf

*Algozzine, B., Wang, C., White, R., Cooke, N., Marr, M. B., Algozzine, K., … Zamora Duran, G. (2012). Effects of multi-tier academic and behavior instruction on difficult-to-teach students. *Exceptional Children*, *79*, 45–64. https://doi.org/10.1177/001440291207900103

Anderson, A. R., Christenson, S. L., Sinclair, M. F., & Lehr, C. A. (2004). Check & Connect: The importance of relationships for promoting engagement with school. *Journal of School Psychology*, *42*(2), 95–113. https://doi.org/10.1016/j.jsp.2004.01.002

*Arnold, K. R. (2012). *The effectiveness of school-wide positive behavior programs in Georgia middle schools* (Doctoral dissertations and Projects). Retrieved from https://digitalcommons.liberty.edu/doctoral/504

Benner, G. J., Nelson, J. R., Sanders, E. A., & Ralston, N. C. (2012). Behavior intervention for students with externalizing behavior problems: Primary-level standard protocol. *Exceptional Children*, *78*(2), 181–198. https://doi.org/10.1177/001440291207800203

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, *1*, 289–328. https://doi.org/10.1080/19345740802400072

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, UK: John Wiley & Sons.

*Bradshaw, C. P., Koth, C. W., Bevans, K. B., Ialongo, N., & Leaf, P. J. (2008). The impact of school-wide positive behavioral interventions and supports (PBIS) on the organizational health of elementary schools. *School Psychology Quarterly*, *23*(4), 462–473. https://doi.org/10.1037/a0012883

*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. https://doi.org/10.1007/s11121-008-0114-9

*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. https://doi.org/10.1177/1098300709334798

*Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on child behavior problems. *Pediatrics*, *130*, e1136–e1145. https://doi.org/10.1542/peds.2012-0243

Caldarella, P., Shatzer, R. H., Gray, K. M., Young, K. R., & Young, E. L. (2011). The effects of school-wide positive behavior support on middle school climate and student outcomes. *RMLE Online*, *35*, 1–14. https://doi.org/10.1080/19404476.2011.11462087

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. https://doi.org/10.1177/10983007070090040301

Cook, B. G., Buysse, V., Klingner, J., Landrum, T. J., McWilliam, R. A., Tankersley, M., & Test, D. W. (2015). CEC's standards for classifying the evidence base of practices in special education. *Remedial and Special Education*, *36*, 220–234. https://doi.org/10.1177/0741932514557271

Council for Exceptional Children (2014). *Council for Exceptional Children standards for evidence-based practices in special education*. Retrieved from http://www.cec.sped.org/~/media/Files/Standards/Evidence%20based%20Practices%20and%20Practice/EBP%20FINAL.pdf

Crone, D. A., Hawken, L. S., & Horner, R. H. (2010). *Responding to problem behavior in schools: The behavior education program*. New York, NY: Guilford Press.

*Darwin, C. (2016). *The relationships among behavior, teacher efficacy, academic achievement and cultural considerations: Accentuating the positive* (Order No. 10119064). Available from ProQuest Dissertations and Theses Global. (1805559075).

*Dirsmith, J. K. (2013). *The impact of school-wide positive behavioral interventions and supports on academic achievement* (Order No. 3557726). Available from ProQuest Dissertations and Theses Global. (1346190745).

Eacho, T. C. (2013). *Violence and disorder, school climate, and PBIS: The relationship among school climate, student outcomes, and the use of positive behavioral interventions and supports* (Order No. 3590612). Available from ProQuest Dissertations and Theses Global. (1432193995).

*Farrington, J. T. (2018). *The Impact of School-Wide Positive Behavior Support on Third Grade Student Achievement as Indicated by Standardized Test Scores* (Order No. 10830221). Available from ProQuest Dissertations and Theses Global. (2075847517).

Retrieved from https://login.lp.hscl.ufl.edu/login?url=https://search.proquest.com/docview/2075847517?accountid=10920

Fisher, Z., Tipton, E., & Zhipeng, H. (2017). robumeta: Robust variance meta-regression. R package (version 2.0). https://CRAN.R-project.org/package=robumeta

*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. https://doi.org/10.1037/spq0000039

Gage, N. A., Cook, B., & Reichow, B. (2017). Publication bias in special education meta-analyses. *Exceptional Children*, *83*, 428–445. https://doi.org/10.1177/0014402917691016

Gage, N. A., Grasley-Boy, N., Lombardo, M., & Anderson, L. (2019). The Effect of School-Wide Positive Behavior Interventions and Supports on Disciplinary Exclusions: A Conceptual Replication. *Behavioral Disorders*, https://doi.org/10.1177/0198742919896305

*Gage, N. A., Grasley-Boy, N., Peshak George, H., Childs, K., & Kincaid, D. (2018). A quasi-experimental design analysis of the effects of school-wide positive behavior interventions and supports on discipline in Florida. *Journal of Positive Behavior Interventions*, *21*, 50–61. https://doi.org/10.1177/1098300718768208

*Gage, N. A., Lee, A., Grasley-Boy, N., & Peshak George, H. (2018). The impact of school-wide positive behavior interventions and supports on school suspensions: A statewide quasi-experimental analysis. *Journal of Positive Behavior Interventions*, *20*, 217–226. https://doi.org/10.1177/1098300718768204

*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. https://doi.org/10.1177/1098300717693556

Gage, N. A., Rose, C. A., & Kramer, D. A. (2019). When prevention is not enough: The distal effect of school-wide positive behavior interventions and support on students' perceptions of bullying. *Behavioral Disorders*, *45*, 29–40. https://doi.org/10.1177/0198742918810761

*Gage, N. A., & Stevens, R. N. (2018). Rigor, replication, and reproducibility: Increasing the relevance of behavioral disorders research. *Education and Treatment of Children*, *41*, 567–588. https://doi.org/10.1353/etc.2018.0029

*Gage, N. A., Sugai, G., Lewis, T. J., & Brzozowy, S. (2015). Academic achievement and school-wide positive behavior supports. *Journal of Disability Policy Studies*, *25*(4), 199–209. https://doi.org/10.1177/1044207313505647

Gage, N. A., Whitford, D. K., & Katsiyannis, A. (2018). A review of schoolwide positive behavior interventions and supports as a framework for reducing disciplinary exclusions. *The Journal of Special Education*, *52*, 142–151. https://doi.org/10.1177/0022466918767847

Good, R. H. & Kaminski, R. A. (Eds.), 2002). Dynamic indicators of basic early literacy skills (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Available from https://dibels.uoregon.edu.

*Grasley-Boy, N. M., Gage, N. A., & Lombardo, M. (2019). Effect of SWPBIS on disciplinary exclusions for students with and without disabilities. *Exceptional Children*, *86*, 25–39. https://doi.org/10.1177/0014402919854196

Grey, J., & Sime, N. (1989). Findings from the national survey of teachers in England and Wales. In Elton, R. (Ed.), *Discipline in schools. Report of the committee of Inquiry chaired by Lord Elton* (pp. 222–250). London, England: Her Majesty's Stationary Office.

*Guardino, D. M. (2013). *The disproportionate use of discipline: An investigation of the potential impact of school-wide positive behavioral interventions and supports* (Order No. 3563580). Available from ProQuest Dissertations and Theses Global. (1399994802).

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*, 60–87. https://doi.org/10.3102/0162373707299706

*Hirschi, C. G. (2015). *A study of school-wide positive behavior support and behavior intervention support teams and their impact on student behavior in six Missouri middle schools* (Order No. 10030341). Available from ProQuest Dissertations and Theses Global. (1774340873).

*Homen, A. M. (2014). *School-wide positive behavior support effects on school climate ratings in diverse elementary schools.* (Order No. 3733470). Available from ProQuest Dissertations and Theses Global; Social Science Premium Collection. (1736050188). Retrieved from https://login.lp.hscl.ufl.edu/login?url=https://search.proquest.com/docview/1736050188?accountid=10920

Horner, R. H., Kincaid, D., Sugai, G., Lewis, T., Eber, L., Barrett, S., … Johnson, N. (2014). Scaling up school-wide positive behavioral interventions and supports: Experiences of seven states with documented success. *Journal of Positive Behavior Interventions*, *16*, 197–208. https://doi.org/10.1177/1098300713503685

Horner, R. H., Sugai, G., & Anderson, C. M. (2010). Examining the evidence base for school-wide positive behavior support. *Focus on Exceptional Children*, *42*, 1–14. Retrieved from. https://login.lp.hscl.ufl.edu/login?url=https://search.proquest.com/docview/808393227?accountid=10920

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. https://doi.org/10.1177/10983007040060010201

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*(3), 133–144.

Hoy, W. K., & Miskel, C. G. (1996). Education administration: Theory, research, and practice (5th ed.). New York: McGraw-Hill.

*Houchens, G. W., Zhang, J., Davis, K., Niu, C., Chon, K. H., & Miller, S. (2017). The impact of positive behavior interventions and supports on teachers' perceptions of teaching conditions and student achievement. *Journal of Positive Behavior Interventions*, *19*, 168–179. https://doi.org/10.1177/1098300717696938

Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2009). Teacher observation of classroom adaptation—checklist: Development and factor structure. *Measurement and Evaluation in Counseling and Development*, *42*, 15–30. https://doi.org/10.1177/0748175609333560

Kraft, M. A. (2018). *Interpreting effect sizes of education interventions*. Brown University Working Paper. Retrieved from https://scholar.harvard.edu/files/mkraft/files/kraft2018interpretingeffectsizes.pdf

Lane, K. L., Wehby, J., Menzies, H. M., Doukas, G. L., Munton, S. M., & Gregg, R. M. (2003). Social skills instruction for students at risk for antisocial behavior: The effects of small-group instruction. *Behavioral Disorders*, *28*, 229–248. https://doi.org/10.1177/019874290302800308

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications, Inc.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., … Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. Washington, DC: U. S. Department of Education, Institute of Education Science, National Center for Special Education Research. Retrieved from https://files.eric.ed.gov/fulltext/ED537446.pdf

McGraw Hill (1992). Comprehensive test of basic skills (5th ed.). Monterey, CA: CTB/Macmillan/McGraw Hill.

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

Mitchell, B. S., Hatton, H., & Lewis, T. J. (2018). An examination of the evidence-base of school-wide positive behavior interventions and supports through two quality appraisal processes. *Journal of Positive Behavior Interventions*, *20*, 239–250. https://doi.org/10.1177/1098300718768217

*Mozley, J. W., Jr. (2016). *The relationship between the implementation of the positive behavioral interventions and supports framework and racial disparities in discipline practices* (Order No. 10168118). Available from ProQuest Dissertations and Theses Global. (1826872933).

*Nelson, J. R., Martella, R. M., & Marchand-Martella, N. (2002). Maximizing student learning the effects of a comprehensive school-based program for preventing problem behaviors. *Journal of Emotional and Behavioral Disorders*, *10*, 136–148. https://doi.org/10.1177/10634266020100030201

Office of Superintendent of Instruction (1996). Washington assessment of student learning, Olympia, WA: Washington State Department of Education.

Pas, E. T., Ryoo, J. H., Musci, R. J., & Bradshaw, C. P. (2019). A state-wide quasi-experimental effectiveness study of the scale-up of school-wide positive behavioral interventions and supports. *Journal of school psychology*, *73*, 41–55. https://doi.org/10.1016/j.jsp.2019.03.001

*Porter, J. A. (2012). *School-wide positive behavior support: Can it help address the problem of disproportionate minority representation in the emotional disturbance disability category?* (Order No. 3546633). Available from ProQuest Dissertations and Theses Global. (1266841438).

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

*Rhodes-Monette, K. (2014). *An analysis of four implementation levels for the positive behavioral interventions and supports (PBIS) on selected aggregated high school student performance outcomes* (Order No. 3717817). Available from ProQuest Dissertations and Theses Global. (1716036569).

Ross, S. W., Romer, N., & Horner, R. H. (2012). Teacher well-being and the implementation of school-wide positive behavior interventions and supports. *Journal of Positive Behavior Interventions*, *14*, 118–128. https://doi.org/10.1177/1098300711413820

*Ryoo, J. H., Hong, S., Bart, W. M., Shin, J., & Bradshaw, C. P. (2018). Investigating the effect of school-wide positive behavioral interventions and supports on student learning and behavioral problems in elementary and middle schools. *Psychology in the Schools*, *55*, 629–643. https://doi.org/10.1002/pits.22134

*Schultz, E. K. (2007). *State and local level implementation of schoolwide positive behavior support: An examination of the Texas behavior support initiative (TBSI)* (Order No. 3276466). Available from ProQuest Dissertations and Theses Global. (304821024).

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*(1), 5–16.

Smolkowski, K. (2006). *An evaluation of school-wide positive behavior support in middle schools with the Oregon healthy teens student survey* (Order No. 3224121). Available from ProQuest Dissertations and Theses Global. (305251887).

Solomon, B. G., Klein, S. A., Hintze, J. M., Cressey, J. M., & Peller, S. L. (2012). A meta-analysis of school-wide positive behavior support: An exploratory study using single-case synthesis. *Psychology in the Schools*, *49*, 105–121. https://doi.org/10.1002/pits.20625

Spokane Public School District (1997). Student safety survey, Spokane, WA: Author.

Sugai, G., & Horner, R. H. (2009). Defining and describing schoolwide positive behavior support. In W. Sailor, G. Dunlap, G. Sugai & R. Horner (Eds.), *Handbook of positive behavior support* (pp. 307–326). New York, NY: Springer US.

Sugai, G., Horner, R. H., Lewis-Palmer, T., & Rossetto Dickey, C. (2012). Team Implementation Checklist 3.1. Adapted from Sugai, Horner, Lewis-Palmer, 2001, *Educational and Community Supports*, *University of Oregon*.

Sugai, G., Horner, R. H., & Todd, A. (2003). Self-assessment survey, Eugene, OR: University of Oregon. https://www.pbisassessment.org/Evaluation/Surveys

*Sørlie, M. A., & Ogden, T. (2015). School-wide positive behavior support–Norway: Impacts on problem behavior and classroom climate. *International Journal of School & Educational Psychology*, *3*, 202–217. https://doi.org/10.1080/21683603.2015.1060912

Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handing complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental Life Course Criminology*, *2*, 85–112. https://doi.org/10.1007/s40865-016-0026-5

*Taylor, R. L. (2017). *The relationship between positive behavioral interventions and supports in elementary schools and mathematics achievement* (Electronic theses and dissertations, University of Louisville; Paper 2876). Retrieved from https://doi.org/10.18297/etd/2876

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48. http://www.jstatsoft.org/v36/i03/

Vincent, C. G., Randall, C., Cartledge, G., Tobin, T. J., & Swain-Bradway, J. (2011). Toward a conceptual integration of cultural responsiveness and schoolwide positive behavior support. *Journal of Positive Behavior Interventions*, *13*, 219–229. https://doi.org/10.1177/1098300711399765

*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*(2), 149–156. https://doi.org/10.1001/archpediatrics.2011.755

Ward, B., & Gersten, R. (2013). A randomized evaluation of the safe and civil schools model for positive behavioral interventions and supports at elementary schools in a large urban school district. *School Psychology Review*, *42*, 317–333. https://login.lp.hscl.ufl.edu/login?, https://search.proquest.com/docview/1507777859?accountid=10920

What Works Clearinghouse (WWC) (2019). What Works Clearinghouse procedures and standards handbook (Version 4.0), Washington, DC: Author. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Wills, H., Kamps, D., Abbott, M., Bannister, H., & Kaufman, J. (2010). Classroom observations and effects of reading interventions for students at risk for emotional and behavioral disorders. *Behavioral Disorders*, *35*, 103–119. https://doi.org/10.1177/019874291003500203

How to cite this article: Lee A, Gage NA. Updating and expanding systematic reviews and meta-analyses on the effects of school-wide positive behavior interventions and supports. *Psychol Schs*. 2020;57:783–804. https://doi.org/10.1002/pits.22336