A META-ANALYSIS OF SCHOOL-WIDE POSITIVE BEHAVIOR SUPPORT: AN EXPLORATORY STUDY USING SINGLE-CASE SYNTHESIS

BENJAMIN G. SOLOMON, SUZANNE A. KLEIN, JOHN M. HINTZE, JAMES M. CRESEY, AND SARAH L. PELLER

University of Massachusetts Amherst

Positive Behavior Support (PBS) for behavioral problems was included in the 1997 Individuals With Disabilities Education Act reauthorization, reflecting the increased implementation and strengthening empirical evidence for PBS in schools. Whereas PBS can be used reactively, its flexibility has led to a popular comprehensive school-wide model used for prevention. School-Wide Positive Behavioral Support (SWPBS) has been used across a variety of school environments and various demographics and has been evaluated using a variety of different outcome measures. The authors conducted a meta-analysis of SWPBS research spanning 16 years and 20 articles. Specifically, single-case studies were evaluated using a regression-based procedure. Results show promising early trends in the data across dependent variables with a need for further research in specific areas.

School-Wide Positive Behavior Support (SWPBS) is an approach to preventing undesirable problem behaviors and promoting positive behaviors that has emerged in recent years as an alternative to more exclusionary and punitive forms of school-wide discipline. Although SWPBS is not a packaged curriculum, nor a scripted intervention, it is an approach that combines evidence-based practices from Positive Behavior Support (PBS; Horner, 1990; Sugai et al., 2000) with the practices of universal behavioral prevention (Walker, Horner, Sugai, & Bullis, 1996) and an instructional focus on teaching positive and prosocial behaviors (Horner et al., 2009; Sugai & Horner, 2009b). SWPBS is growing widely in school-based practice in the United States, with more than 30 states reporting the establishment of state-wide SWPBS leadership teams and more than 7,900 schools reporting the adoption of SWPBS as a school-wide prevention and intervention system (Spaulding, Horner, May, & Vincent, 2008). SWPBS has also been incorporated into the movement toward Response to Intervention (RTI) and a three-tiered model of prevention and intervention for academic and behavioral supports (Sailor, Doolittle, Bradley, & Danielson, 2009; Sugai & Horner, 2009).

In response to the current boom of interest in the SWPBS approach among educators (Spaulding et al. [2008] predicted that federal and state requirements would fuel a further increase in implementation), a meta-analysis of existing research would provide helpful information to researchers and practitioners as they continue to develop and monitor the effectiveness of SWPBS. Meta-analyses of PBS exist at the individual and small group level (Carr et al., 1999; Marquis et al., 2000). No meta-analytic synthesis of research on SWPBS has been conducted, however. One reason for this may be the methodological nature of the majority of published research on SWPBS – primarily single-case designs. Although meta-analytic procedures with group design approaches are firmly established (Glass, McGaw, & Smith, 1981), the quantitative synthesis of single-case data is less understood.

PBS

Before SWPBS was developed as a preventive approach, PBS for individuals with disabilities was an intervention that set the stage for its current adaptation into a more comprehensive and preventative model. Key points from this history include the need that emerged in the 1980s for

We thank Richard Parker and Andrew Koerner for their guidance on methods of single-case statistical analysis. Correspondence to: Benjamin Solomon, c/o John Hintze, 362 Hills House South, University of Massachusetts Amherst, Amherst, MA 01003. E-mail: benjamin.solomon@okstate.edu
non-aversive behavioral strategies in the treatment of individuals with severe disabilities. In the late 1980s, federal funding was allocated for research and development for the pursuit of a “technology of non-aversive behavioral support,” which was eventually termed “positive behavioral support” by Horner (1990). The principles of applied behavioral analysis were used as a foundation for developing PBS, which focuses on positive reinforcement to support a student’s performance of socially desirable target behaviors. PBS emphasizes that behavioral modification that focuses on aversive and exclusionary discipline strategies with no focus on increasing the frequency of positive behavior is usually ineffective and possibly inhumane.

In response to the success of PBS with students across a range of special education categories, the 1997 reauthorization of the Individuals with Disabilities Education Act (IDEA) included new language requiring the use of “positive behavioral intervention strategies and supports” for any child in special education with emotional and behavioral problems (IDEA, 1997). The following year, in response to this new call for more formal and widespread use of PBS, the U.S. Office of Special Education Programs created an online technical assistance center with resources for educators and administrators who were implementing PBS (Sugai et al., 2000). The first efforts to develop and implement SWPBS began in the late 1980s and early 1990s (Horner & Sugai, 2007; Walker et al., 1996).

SWPBS

SWPBS is still a developing model of prevention and intervention that continues to be refined through research. Sugai and Horner (2009b) stated that “SWPBS is not a curriculum, intervention, or program. However, it is an approach designed to improve the adoption, accurate implementation, and sustained use of evidence-based practices related to behavior and classroom management and school discipline systems” (p. 309). Although specific application of SWPBS varies by school, five common core components serve as the foundation. Behavioral theory and applied behavioral analysis are the first and earliest influences of SWPBS. The use of positive reinforcement and functional behavioral assessment are the strongest underlying influences of SWPBS in practice. Second, the focus on prevention is a key feature that distinguishes SWPBS from individually applied PBS. Third, an instructional focus permeates the interventions and behavioral teaching practices that comprise SWPBS. Fourth, SWPBS draws from evidence-based behavioral practices to increase the likelihood of interventions being effective and generalizable. Last, a systems approach is a defining feature of SWPBS, making use of existing school resources and structures to infuse the culture and practices of the school system with the SWPBS approach (Sugai & Horner, 2009b). Given these theoretical components of SWPBS, we describe the key features of SWPBS as it is typically implemented in practice.

Key Features of SWPBS

In SWPBS, rules for student behavior are made explicit, simple, and consistent. Three to five core expectations are chosen for the entire school. An example from one of the studies in the present meta-analysis is presented to illustrate this intervention component. McCurdy, Manella, and Eldridge (2003) listed “Be Responsible, Be Respectful, Be Ready” as the core expectations of an urban elementary school that used SWPBS to reduce disruptive and antisocial behavior. Each of these three expectations was explicitly defined in target behaviors for each environment of the school, such as the cafeteria, classroom, hallways, and playground. Specific behaviors such as “Use a quiet voice at all times” were phrased in the positive voice, rather than the use of “do not” phrasing (e.g., “do not talk in a loud voice”).
To illustrate the instructional focus of SWPBS, this example school is discussed further. Teachers in the school reported by McCurdy et al. (2003) planned behavioral lessons before the beginning of the school year and booster sessions at key times throughout the year. Teachers taught the expectations and target behaviors to their students in classroom settings as well as in other target environments of the school.

SWPBS provides acknowledgment for successfully meeting behavioral expectations. Acknowledgment systems can be similar to a traditional token economy historically used in behavioral intervention systems. In addition to tokens or tickets, acknowledgment is also provided to emphasize positive social attention from teachers as an important prosocial source of positive reinforcement. “Top Dawg” tickets and “T.N.T.” (teachers noticing talent) tickets are two examples of acknowledgment systems that have been used in middle schools (Metzler, Biglan, Rusby, & Sprague, 2001). Students are then able to exchange their tickets to purchase prizes or to participate in a school raffle.

When SWPBS is implemented in a school, the existing policies for office discipline referrals (ODRs), detentions, and suspensions may still be kept as part of the system. Teachers, however, are encouraged in their use of immediate corrective feedback after a behavioral problem has occurred and are encouraged to respond to behavioral problems in a similar fashion as they respond to academic problems: with correction and teaching.

As PBS has expanded from use with individual students to a school-wide model of prevention and intervention, it has often been incorporated into a three-tiered model of service delivery. The three-tiered model was adapted from the field of public health (Institute of Medicine, 1994) and uses a population-based framework for providing both academic and behavioral prevention and intervention programming. In a school using SWPBS, all students are served at the primary prevention level with universal programming and interventions. At the secondary prevention level, specialized group systems are provided for students exhibiting mild to moderate levels of behavior problems. Finally, students presenting with the most severe or high-risk behavior problems are provided with tertiary preventative support at an individualized level (Sugai, 2007). Thus each component of SWPBS, such as the explicit teaching of target behaviors or the acknowledgment and reinforcement system, can be presented in various levels of frequency, duration, and intensity.

Another important feature of SWPBS is the use of data collection to inform decisions about how to meet the needs of students along the three-tier model. The two most common systems for collecting ongoing data related to SWPBS implementation include the School-Wide Evaluation Tool (SET; Horner et al., 2004) and the School-Wide Information System (Educational and Community Supports, 2007). The SET is used to measure the treatment integrity of SWPBS practices in teachers and staff. The School-Wide Information System is an online database that is used to record ODRs, suspensions, and other office records of student conduct problems. After being collected, data can be summarized detailed by student, by grade level, by referring teacher, by location in the school, by type of infraction, by time of day, and by time of year (e.g., Irvin, Tobin, Sprague, Sugai, & Vincent, 2004). Data summaries are then used by an SWPBS leadership team for formative intervention planning (e.g., increase hallway supervision after meals, alter the bus dismissal routine to improve student behavior).

History of Embedded Practices

Programs that are currently recognized as SWPBS have operated under different terms throughout the last two decades. In 1993, Colvin, Kame’enui, and Sugai introduced Project PREPARE. This early universal behavioral prevention model focused on (a) a consistent approach to managing behavioral problems, (b) viewing school discipline as a prerequisite of academic learning, (c) a focus on proactive prevention, (d) active involvement from school administration, (e) faculty...
Solomon et al.

commitment, and (f) a focus on effective staff development. The first single-case study was by Colvin, Sugai, Good, and Lee (1997), who began a line of research on core components of school-wide behavior intervention in an elementary school in the Northwest. They implemented a school-wide “intervention plan” consisting of consultation on teachers’ “precorrection” and “active supervision.” They measured effects on student behavior during transitions into and out of the building and to and from lunch. The following year, a similar program’s effects were examined on behavior during recess and hallway transitions and in the cafeteria (Lewis, Sugai, & Colvin, 1998). The school-wide system in place at the school was not yet termed “Positive Behavior Support;” these authors refer to the program as a “social skill instruction program combined with direct intervention on problem behavior” (Lewis et al., 1998, p. 446). As in later school-wide systems under the name SWPBS, these researchers began by establishing a set of positively stated school rules and initiating a school-wide token reinforcement system.

“Pre-correction” is defined in the research as “presenting an instructional prompt prior to a student entering a context in which a problem behavior is likely to occur” (DePry & Sugai, 2002, p. 257). “Active supervision” is operationalized to consist of “circulating around the classroom, scanning the classroom, interacting with students, and reinforcing demonstrations of expected academic and social behaviors” (p. 257). The acknowledgment system that has become a hallmark of SWPBS implementation can be described as a more positive, consistent, and proactive application of group contingency reinforcement principles, which have a long history of use in schools and have substantial support in the school psychology and applied behavior research (Kelshaw-Levering, Sterling-Turner, Henry, & Skinner, 2000; Stage & Quiroz, 1997; Tankersley, 1995).

In 2000, Sugai et al. identified systems-level PBS as “the application of positive behavioral intervention and systems to achieve socially important behavior change” (p. 133). In 2002, Depry and Sugai published an account of a classroom intervention combining consultation around active supervision, precorrection, and work on teachers’ “antecedent prompts” within the school-wide system. By this time, however, other researchers were beginning to adopt the term “school-wide positive behavior supports” in recognition of the increasing use of PBS principles in primary prevention. One group of researchers, who used PBS strategies to reduce problem behavior during recess periods, defined the emerging universal system as a “continuum of positive behavior supports,” including interventions at the preventive, early intervention, and individualized tiers (Lewis, Powers, Kelk, & Newcomer, 2002).

Quantitative Analysis of Single-Case Data

Meta-analysis is used to evaluate the generalizability of a pattern of responses across studies to an independent variable (IV). Meta-analysis has evolved rapidly over the past four decades, now being a popular, scientifically defensible induction technique encompassing a wide range of methods. A literature review of even a few studies can benefit from adopting accepted statistical techniques to ward against bias (Glass, 1977). Meta-analysis has traditionally catered to group designs and correlational research. Single-case data have historically relied on either visual analysis or methods that do not align with normal statistical assumptions (e.g., percentage of non-overlapping data, comparison of level changes), resulting in an absence of quantitative data in the published literature. Statistically defensible methods to create and synthesize single-case data and the requisite technology has existed, however, for more than a decade.

Purpose of the Current Study

To date, a number of published studies have provided positive support for SWPBS on improving student behavior at both the individual and school levels (Chorpita, 2008; Drake, Latimer, Leff,
Although individual studies can provide excellent information regarding efficacy and effectiveness of local applications, at a broader level the generalizability of the findings is restricted. This limitation is compounded by the broad flexibility of SWPBS and its operation under different labels in the literature. As of yet, no research that we found attempted to review the various extensions of SWPBS in a quantitative fashion. Such an analysis could provide valuable insight into where future research efforts should be directed as well as inform practitioners how confident they can be in implementing SWPBS in different contexts. The current study focused on examining the effects of SWPBS across different school environments, time frames, and outcome variables.

The purpose of the current study was to apply meta-analytic procedures to single-case dependent variables (DVs) typically used to quantify change in student behavior in response to SWPBS. Results are presented along the continuum of several outcome metrics commonly used with SWPBS, including setting and duration of the study.

**METHOD**

**Criteria for Study Inclusion**

Either studies published in peer-reviewed journals or published dissertations were included in the current analysis. Online searches were conducted in May 2008. The PsycINFO database was searched using the keywords *positive behavior support* and *school-wide* yielding 24 articles. Changing the keyword to *whole school* yielded an additional two studies. Searches within the Educational Resources Information Center (ERIC) database using the same keywords yielded 26 studies. Dissertation abstracts were searched using the keywords, which yielded 41 studies. Additionally, 53 articles were located through a bibliography on the Positive Behavior Interventions and Support website (Horner & Sugai, 2007). Multiple databases were searched in an effort to uncover every relevant article; however, many of those found were duplicates between the databases. Further studies were located through a bibliographic search of the studies indicated above. The online searches were replicated in June 2009, yielding one more article. Studies' publication from 1993 to 2008.

Studies were screened using the following criteria for inclusion:

1. The study examined “Positive Behavior Support” by both explicitly defining and examining at least one individual component (as previously described).
2. The study used PBS on a school-wide level rather than with individual students. This can also be understood as application of PBS at the universal level.
3. The study’s primary focus was change in student behavior over time in a school setting as a result of implementing SWPBS.
4. The study used a single-case design. Expert opinion on single-case analysis suggests that it is not methodologically sound to include both group and single-case designs within the same meta-analysis, as effect size values are largely a function of the design used (Faith, Allison, & Gorman, 1996).
5. Data from each study were displayed in a graphic format (e.g., a time-series display) so that data could be digitized for analysis purposes and a sufficient number of datum points were available in order to calculate a regression line (at least three baseline datum points and three intervention datum points. A description of how low \( N \) was controlled for is presented later in this article). Single-case studies rarely provide inferential statistics to complement analysis, and the bulk of necessary data can be extracted only from the graph(s) typically included.
6. The study was published in English.

After limiting the articles to be included in the analysis based on these criteria, 20 studies were identified that included effects of SWPBS components or a comprehensive SWPBS program. Eighteen of the studies were published in peer-reviewed journals between 1996 and 2008, and
<table>
<thead>
<tr>
<th>Category</th>
<th>No. of studies</th>
<th>No. of effect sizes</th>
<th>Mean effect size</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>DV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ODRs</td>
<td>7</td>
<td>8</td>
<td>.33</td>
<td>.22–.45</td>
</tr>
<tr>
<td>Problem behavior</td>
<td>10</td>
<td>10</td>
<td>.44</td>
<td>.38–.49</td>
</tr>
<tr>
<td>Setting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstructured</td>
<td>10</td>
<td>10</td>
<td>.51</td>
<td>.45–.63</td>
</tr>
<tr>
<td>Classroom</td>
<td>8</td>
<td>9</td>
<td>.27</td>
<td>.18–.36</td>
</tr>
<tr>
<td>Duration of intervention</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 1 year</td>
<td>12</td>
<td>12</td>
<td>.47</td>
<td>.41–.52</td>
</tr>
<tr>
<td>1 year to 2 years</td>
<td>8</td>
<td>9</td>
<td>.50</td>
<td>.37–.61</td>
</tr>
<tr>
<td>Type of intervention</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensive</td>
<td>12</td>
<td>13</td>
<td>.48</td>
<td>.42–.54</td>
</tr>
<tr>
<td>Grade level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary (K–5)</td>
<td>13</td>
<td>14</td>
<td>.51</td>
<td>.45–.57</td>
</tr>
<tr>
<td>Middle (6–8)</td>
<td>3</td>
<td>3</td>
<td>.60</td>
<td>.49–.69</td>
</tr>
<tr>
<td>Demographic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>6</td>
<td>7</td>
<td>.56</td>
<td>.48–.64</td>
</tr>
<tr>
<td>Rural</td>
<td>5</td>
<td>5</td>
<td>.45</td>
<td>.34–.56</td>
</tr>
<tr>
<td>Suburban</td>
<td>3</td>
<td>3</td>
<td>.46</td>
<td>.34–.57</td>
</tr>
</tbody>
</table>

*Note.* To avoid dependency, each study could be used only once in each category.

Two were published dissertations. Twelve of the studies measured effects of a multicomponent implementation (what one would consider to be a full SWPBS program), whereas eight examined a smaller number of components implemented in various combinations. For this study, SWPBS is described as comprehensive when it includes all five of the features described or component when it fails to address them all.

Specific categories of analysis are described in Table 1. DV used constituted the first category. The most common DV was the frequency of incidents of problem behavior, based on some type of structured direct behavior observation. Ten of the studies measured SWPBS effects using this DV. The second most common DV was the schools’ incidence of ODRs. ODRs are important because they represent a measure of (a) how prepared teachers are to deal with minor behavior problems within the classroom without major disruption to a lesson and (b) the actual prevention of problem behaviors. ODRs, because of convenience and frequency, are a common DV in SWPBS, both as an evaluative tool of an SWPBS model and for data-tracking purposes of individual students (Educational and Community Supports, 2011; Irvin et al., 2004). Seven of the studies used ODRs to measure effects. Many studies measured effects on several different DVs simultaneously. Each study, however, could contribute only one effect size to each category to prevent dependency from becoming a confound. Interestingly, Kartub, Taylor-Greene, March, and Horner (2000) measured (in decibels) the effects of the full program on hallway noise.

Setting of the intervention was coded to form another category. Two school-based settings were examined: unstructured spaces, such as the hallway or recess areas (10 studies), and the classroom (eight studies). In addition, the duration of the intervention phase was coded, varying from less than a year (12 studies) to one to two years of implementation (eight studies). Studies that combined all five features of SWPBS were labeled “Comprehensive” (12 studies). Other studies used select strategies at the school level, such as active supervision or pre-teaching procedures, but were not aggregated into a single category because of the heterogeneity of the combinations.

*Psychology in the Schools*  DOI: 10.1002/pits
were coded as taking place at urban (six studies), rural (five studies), or suburban locations (three studies). Studies that offered no information about the demographics of the participant(s) or identified multiple demographics were not included in this category. School level was also reviewed. Of the 20 studies, 13 were conducted in elementary schools with the balance in preschools, middle schools, and high schools.

Procedures

Method for analyzing research articles. Line and bar graphs from each study were extracted from hard copy articles or .pdf files and saved as .gif files. The cartography program XYit (Geomatix, 2005) was then used to extract the data from each graph. Data were recalculated by assigning a relative Y value to each data point after the boundaries of the graph had been calibrated (Parker & Brossart, 2003; Parker et al., 2005). After being calculated, all data were subsequently imported into Microsoft Office Excel 2007.

Effect size calculations. Because of issues related to autocorrelation and lack of power associated with simple mean difference approaches, a regression-based procedure was used to calculate effect sizes (Faith et al., 1996; Parker et al., 2005). The advantage of such an approach is that it controls for autocorrelation while maintaining an adequate level of power (Faith et al., 1996). Specifically, the current study used the Allison-MT method (Faith et al., 1996). In comparison to other procedures for calculating single-case effect sizes, the Allison-MT procedure has a high level of power while at the same time yielding virtually no autocorrelation (Parker et al., 2005). The Allison-MT method accounts for both trend and mean differences between phases and can be summarized as follows: (a) A line of best fit is calculated based on baseline data; (b) residuals for intervention data are calculated by comparing observed to predicted data on the basis of what would be predicted by baseline trend; (c) data are dummy coded by phase to create a new variable; and (d) residuals are regressed onto the dummy-coded variable.

Last, results were converted to logits for further calculations (e.g., averaging), and then finally converted to $R^2$ proportion of variance effect sizes for reporting purposes (Parker & Brossart, 2003). Whereas effect sizes calculated using regression always result in a positive value, the effect size was assigned a negative value if the outcome variable showed a decrease in the direction opposite the stated research hypothesis. This reassignment of value, however, applied to only one DV of one study (i.e., Blevins, 2003).

The procedure just described was conducted for each line or bar graph presented in included studies. For multiple baseline studies, a separate effect size was calculated for each setting or participant and then averaged. Studies were then weighted by number of observational points using the inverse variance method to provide more influence to studies with a larger sample size – the equivalent of controlling for sample size in a group study (Faith et al., 1996). Studies were then grouped into categories. Within each subcategory, each complete design received its own weight. For instance, a multiple baseline design across three participants was not counted three times, but rather the effect sizes were averaged and the baseline and intervention points summed to be used in the weighting procedure. A study that had two single-case designs that were independent of each other with regard to participant and setting were weighted and counted separately.

Confidence intervals (CIs) were calculated by using the standard error of the weighted mean effect size. This standard error is equivalent to the square root of the inverse of the sum of calculated

1 While detrending only the baseline data can create instability with studies with low baseline $N$, Faith et al. (1996) make the logical point that this is more a weakness in the study than in the procedure.
2 If the correlations of level to detrended data and slope to detrended data are of opposite signs, the raw data, instead of the residuals, are regressed onto the dummy-coded variable (Faith et al., 1996).
variances for each subcategory. Organizing the results in this manner was done instead of calculating an overall SWPBS effect size because a primary goal of the study was to examine under what specific contexts SWPBS was most effective, what measures are most reliable, and where future research needs to occur. CIs can be seen as a type of significance testing. If the CI of any given variable crosses the zero threshold, results are not statistically significant. Additionally, CIs within categories can be compared as a form of moderation analysis.

**Reliability of data collection.** To control for error associated with the digitizing procedure, six randomly selected graphs were digitized by the second author (initial digitizing was conducted by the first author). Across the sampled graphs, an average correlation between authors was $r = .98$.

**Results**

Individual effect sizes from each included study, the specific population used, the type of design, the duration of the study, the specific intervention implemented, and the outcome measures used to analyze change are presented in Table 2. Overall, 21 effect sizes were calculated across 20 studies. Mean values across categories were moderate and positive in size ranging from $r^2 = .27$ to $r^2 = .60$ as a result of SWPBS implementation. Categorical results are presented as 95% CIs in Figure 1. Within categories (e.g., DV, demographic), any overlap between CIs suggests that the category does not significantly moderate the effectiveness of SWPBS. Alternatively, CIs that do not overlap within category suggest significant differences. Moderation was demonstrated only for the setting of implementation. SWPBS in unstructured spaces was significantly more effective [.44, .63] than in the classroom [.18, .36]. Research has demonstrated that single-case $R^2$ effect sizes do not meet traditional markers set forth by Cohen (low = .01, medium = .06, high = .14; Parker

![Figure 1. Weighted means and CI’s of effect sizes. X-axis ticks align with Parker et al.’s (2005) Allison-MT distribution quartiles.](image-url)
Table 2

Attributes and Effect Sizes of Individual Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Subjects</th>
<th>Design</th>
<th>Duration</th>
<th>Intervention</th>
<th>Outcome measures</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benedict, Horner, &amp; Squires (2007)</td>
<td>4 Head Start classrooms</td>
<td>Multiple baseline across subject</td>
<td>&gt;1 year</td>
<td>SWPBS consultation</td>
<td>% of Problem behavior Teacher fidelity</td>
<td>.12</td>
</tr>
<tr>
<td>Blevins (2007)</td>
<td>Rural high school</td>
<td>AB</td>
<td>1–2 years</td>
<td>SWPBS consultation</td>
<td>ODRs Suspension Expulsion Attendance Teacher Fidelity</td>
<td>&gt; .01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colvin, Sugai, Good, &amp; Lee (1997)</td>
<td>Rural/Suburban elementary school transition Times</td>
<td>Multiple baseline across setting</td>
<td>&gt;1 year</td>
<td>Active supervision and pre-correction</td>
<td>Frequency of Problem Behavior Staff Interactions</td>
<td>.55</td>
</tr>
<tr>
<td>Depry &amp; Sugai (2002)</td>
<td>6th grade classroom</td>
<td>ABAB</td>
<td>&gt;1 year</td>
<td>Active supervision and pre-correction</td>
<td>Frequency of Minor Behavioral incidents.</td>
<td>.02</td>
</tr>
<tr>
<td>Franzen &amp; Kamps (2008)</td>
<td>Urban elementary school recess</td>
<td>Multiple baseline across setting</td>
<td>&gt;1 year</td>
<td>Active supervision</td>
<td>Frequency of problem Behavior Teacher fidelity</td>
<td>.68</td>
</tr>
<tr>
<td>Johnson-Gros, Lyons, &amp; Griffin, 2008</td>
<td>Rural high school hallway</td>
<td>Multiple baseline across Setting</td>
<td>&gt;1 year</td>
<td>Active supervision</td>
<td>ODRs</td>
<td>.09</td>
</tr>
<tr>
<td>Kartub, Taylor, March, &amp; Horner (2000)</td>
<td>Rural middle School Hallway</td>
<td>AB</td>
<td>&gt;1 year</td>
<td>SWPBS consultation</td>
<td>Decibel level</td>
<td>.56</td>
</tr>
<tr>
<td>Lewis, Colvin, &amp; Sugai (2000)</td>
<td>Suburban/Rural elementary school recess groups</td>
<td>Multiple baseline across setting</td>
<td>&gt;1 year</td>
<td>Active supervision and pre-correction</td>
<td>Rate of problem Behavior Under: Structured activities Unstructured activities Teacher fidelity</td>
<td>.09</td>
</tr>
<tr>
<td>Lewis, Powers, Kelk, &amp; Newcomer (2002)</td>
<td>Elementary school recess Groups</td>
<td>Multiple baseline across Setting</td>
<td>&gt;1 year</td>
<td>Group contingency and pre-correction</td>
<td>Frequency of Problem Behavior</td>
<td>.49</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Study</th>
<th>Subjects</th>
<th>Design</th>
<th>Duration</th>
<th>Intervention</th>
<th>Outcome measures</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lewis, Sugai, &amp; Colvin (1998)</td>
<td>Elementary school transition time, recess and cafeteria</td>
<td>Multiple baseline across setting</td>
<td>&gt;1 year</td>
<td>Contingency and social skills instruction</td>
<td>Rate of problem Behavior</td>
<td>.47</td>
</tr>
<tr>
<td>Luiselli, Putnam, Handler, &amp; Feinberg (2005)</td>
<td>Urban elementary school</td>
<td>AB</td>
<td>1–2 years</td>
<td>SWPBS consultation</td>
<td>ODRs Suspensions, Reading/Math achievement, Teacher feedback</td>
<td>.21, .32</td>
</tr>
<tr>
<td>McCart (2003)</td>
<td>Urban elementary school</td>
<td>AB</td>
<td>&gt;1 year</td>
<td>SWPBS consultation</td>
<td>ODRs (School 1), ODRs (School 2)</td>
<td>.13, .16</td>
</tr>
<tr>
<td>McCurdy, Mannella, &amp; Eldridge (2003)</td>
<td>Urban elementary school</td>
<td>AB</td>
<td>1–2 years</td>
<td>SWPBS consultation</td>
<td>ODRs, Teacher fidelity, Teacher satisfaction</td>
<td>.61, -</td>
</tr>
<tr>
<td>Metzler, Biglan, Rusby, &amp; Sprague (2001)</td>
<td>6th–8th grade</td>
<td>AB</td>
<td>1–2 years</td>
<td>SWPBS consultation</td>
<td>7th Grade ODRs, ODRs males, ODRs females, Student feedback on Safety 6th, Student feedback on Safety 7th, Student feedback on Abuse 6th, Student Feedback on Abuse 7th, Teacher Fidelity 6th, Teacher Fidelity 7th</td>
<td>.73, .24, .51, .51, .72, .71, .64, .01</td>
</tr>
<tr>
<td>Morrison &amp; Jones (2007)</td>
<td>Urban 3rd grade classrooms</td>
<td>Multiple baseline across Setting</td>
<td>&gt;1 year</td>
<td>Positive peer praise</td>
<td>Average weekly Problem Behavior</td>
<td>.44</td>
</tr>
<tr>
<td>Nelson, Colvin, &amp; Smith (1996)</td>
<td>Elementary school breakfast/before School</td>
<td>Multiple baseline across setting</td>
<td>&gt;1 year</td>
<td>SWPBS consultation</td>
<td>% of Positive Behavior, % of Disruptive Behavior</td>
<td>.39, .38</td>
</tr>
<tr>
<td>Putnam, Handler, Ramirez-Platt, &amp; Luiselli (2003)</td>
<td>Urban elementary school buses</td>
<td>ABAB</td>
<td>1–2 years</td>
<td>SWPBS consultation</td>
<td>ODRs, Suspensions</td>
<td>.11, .46</td>
</tr>
</tbody>
</table>
Table 2  
Continued

<table>
<thead>
<tr>
<th>Study</th>
<th>Subjects</th>
<th>Design</th>
<th>Duration</th>
<th>Intervention</th>
<th>Outcome measures</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scott, 2001</td>
<td>Urban elementary school</td>
<td>AB</td>
<td>1–2 years</td>
<td>SWPBS consultation</td>
<td>Hours spent in Discipline room</td>
<td>.78</td>
</tr>
<tr>
<td>Stormont, Smith, &amp; Lewis (2007)</td>
<td>5 Head Start teachers</td>
<td>Multiple baseline across participants</td>
<td>&gt;1 year</td>
<td>Increased praise and pre-correction</td>
<td>Frequency of Praise Frequency of reprimands Frequency of problem Behavior</td>
<td>.65 .44 .54</td>
</tr>
<tr>
<td>Taylor-Greene et al. (1997)</td>
<td>Rural middle school</td>
<td>AB</td>
<td>1–2 years</td>
<td>SWPBS consultation</td>
<td>ODRs</td>
<td>.88</td>
</tr>
</tbody>
</table>

Note. In the event no effect size is presented, the data from this variable could not be extracted or were not measured using single-case methodology.

et al., 2005). They may be evaluated for strength by using the following quartiles as a benchmark: 1st quartile equal to $R^2 = .35$, 2nd quartile equal to $R^2 = .65$, and 3rd quartile equal to $R^2 = .90$ (rounded from the original report; Parker et al., 2005). Overall, effect sizes in this study in comparison to these benchmarks were low to medium.

It is important to mention the implications of using both experimental and quasi-experimental studies in this synthesis. In the present study, experimental research is defined as a single-case design that included at least one full phase reversal or included multiple baselines with appropriate staggering of phase changes. Quasi-experimental research is defined as any study that had no phase reversals (i.e., AB) or did not stagger phase changes when multiple baselines were present. Research embedded within schools often uses quasi-experimental designs to assess program effectiveness. This may be appropriate by increasing external validity, feasibility, and reducing cost of assessment, but it sacrifices internal validity. For example, in implementing SWPBS, it may be extremely difficult to coordinate a phase reversal in a highly populated and scheduled school. Overall, 10 of the analyzed graphs were experimental, and 11 were quasi-experimental. The potential confound of type of design was tested using analogue multiple regression analysis for effect sizes (Hedges, 1994). Effect sizes were treated as the DV and study type was dummy coded and treated as the IV. The regression was weighted using the inverse variances previously described. The resulting CI, $\beta_1 = .10$, 95% CI $[-.03, .23]$, included zero. Therefore, the null hypothesis that effect sizes were equal across study types was not rejected.

Treatment fidelity is a critical component of any intervention plan (Gresham, 1989, 2009), particularly true with regard to SWPBS, which relies largely on a consultee-based approach (Caplan, 1970). Professionals intervene with faculty, who then must implement the behavioral intervention as intended by the consultants. Professional consultation results in two levels of treatment fidelity: the fidelity of implementation of the consultant and of the teacher. Unfortunately, only one study included in the analysis reported consultant-level fidelity: Benedict, Horner, and Squires (2007) recorded the number of hours and sessions spent in consultation for all teachers. Twelve other studies measured teacher-level integrity through direct observation, surveying (such as use of the SET), or implementation checklists. Of these studies, four reported integrity in such a way that an effect size could be generated. The weighted mean effect size for changes in teaching behavior from baseline to intervention was $r^2 = .40$, 95% CI [.27, .57].
Finally, a fail-safe $N$ was calculated to test the extent of publication bias in the current study. Using Orwin’s method (Hedges & Olkin, 1985) with the criterion effect size set to $R^2 = .35$ (the 1st quartile in Parker et al.’s [2005] distribution), it was found that six additional studies would be needed to reduce the averaged results to $R^2 = .35$.

**DISCUSSION**

This study aggregated the published data on SWPBS and critically analyzed results for the purpose of increasing generalizability of findings and providing a practical review to orient future researchers and practitioners. Results were divided into type of DV, setting of implementation, duration of study, grade level, and demographic. Results were promising; however, there was variability across specific subcategories that merits further discussion. It is important to note before the discussion of results that SWPBS is a universal level prevention effort. Individual subjects were typically whole classrooms or even entire schools. As such, it is expected that effect sizes are lower in comparison to the distribution of Parker et al. (2005).

SWPBS’s effect on problem behavior was in the low average range, as measured by direct observation. This finding demonstrates that SWPBS is moderately effective in reducing problem behavior in students. Articles included in this analysis largely did not specify whether problem behavior was observed from students of varying tiers or just at Tier I. At Tier III, when an individualized behavior plan or consideration of a new placement occurs, PBS for individual students has robust evidence (Carr et al., 1999; Marquis et al., 2000). This meta-analysis combined with other reviews shows that PBS is effective at multiple tiers of service in reducing undesirable student behavior.

This finding also highlights the utility of using direct observation as a primary gauge of effectiveness. Practitioners and researchers alike should consider using observation of student problem behavior at a class-wide level when reviewing SWPBS implementation, as direct observation is sensitive to student-level changes associated with SWPBS. ODRs had a nonsignificantly lower average effect size and more variability about the mean. This result is to be expected as ODRs are inherently less reliable than is direct observation. ODRs are indirect measures of student behavior, and they have irregular frequencies of occurrence. There is significant variability from teacher to teacher and from school to school as to when an ODR occurs.

Despite these limitations of ODRs, they are easily measured in schools, promulgating their use. They have been evaluated for technical adequacy and have shown to be a robust measure of student behavior. McIntosh, Cambell, Carter, and Zumbo (2009) established concurrent validity between ODRs and the *Behavior Assessment System for Children* (2nd ed.) – Teacher Rating Form (BASC-2; Reynolds & Kamphaus, 2006). Bivariate correlations between the two measures were strong. Univariate significance testing demonstrated that ODRs effectively separated low- from high-risk students when risk was measured with the BASC-2 Externalizing Composite, Adaptive Composite, and frequency of suspensions. No significant relationship was found between ODRs and the BASC-2 Internalizing Composite. This finding demonstrates that ODRs do have technical adequacy. When possible, ODR data should be supplemented with the more sensitive direct observation of problem behavior to increase the validity of outcome data.

With respect to setting, interventions located at recess, in hallways, and in the cafeteria (unstructured settings) yielded average effects. Active supervision, pre-correction strategies, and group contingency plans were the most popular choices for interventions in these unstructured environments. Interestingly, the effect of SWPBS in the classroom had a significantly lower mean effect size than in unstructured settings. The probability of problem behaviors is greater in unstructured settings. More students inhabit these areas, with less supervision and less consistency of behavioral expectations. Supervisors of unstructured areas may have less training in classroom management. This situation creates a rich potential for SWPBS to swiftly increase positive behavior in these contexts.
This finding is important for practitioners as they consider what school-wide approach to adopt and where to strategically implement it. Behavioral prevention does not necessarily have to be focused exclusively on the classroom, nor should it. Putnam, Handler, Ramirez-Platt, and Luiselli (2003) demonstrated how this notion can be extended to school buses. Buses are typically not well suited to behavioral prevention, despite the contextual features that render buses vulnerable to misbehavior (e.g., little supervision, no structure, high density of students, driver possibly having little behavioral management training). The results of Putnam et al. (2003) highlight how SWPBS can be used effectively when other resources are scarce. This finding also highlights the need for a comprehensive needs assessment when implementing SWPBS. Strategic execution can target areas found to be at risk for behavioral problems, maximizing the efficient use of resources.

Categorizing studies according to the duration of the SWPBS intervention and comparing these subcategories showed that all durations yielded an average effect size of at least $r^2 = .468$. Changes across intervention length were small, suggesting that implementation effects remained fairly constant across time. It is worth noting the paucity of research of multiyear longitudinal designs. Future research would add value to the evidence base of SWPBS by conducting intervention phases of at least two full school years. Extending the length of study is particularly important because it is recommended that a leadership plan for SWPBS implementation take two to three years (Sugai et al., 2010). The goal of SWPBS is to shift community beliefs and behavior to allow permanent sustainability of SWPBS. On the basis of this recommendation, one might not expect SWPBS to reach maximum integrity until three to five years of implementation. No study reviewed implemented SWPBS for more than two years, however. These relatively short durations of intervention beg the question, “How sustainable is SWPBS over time?”

SWPBS has been evaluated at the pre-school, elementary school, middle school, and high school levels. Unfortunately, only two studies from high school and two from pre-K were found in a search of the literature, prohibiting the calculation of CIs. Blevins (2007) and Johnson-Gros, Lyons, and Griffin (2008) found modest results for high school, and Stormont, Smith, and Lewis (2007) and Benedict, Horner, and Squires (2007) found promising results for Head Start classrooms (see Table 2). These findings highlight the need for more research in these areas, particularly as national SWPBS implementation includes more early childhood and secondary settings. The current findings suggest strong effects for both middle school and elementary school settings.

Although not a significant difference, SWPBS in middle school had a higher mean effect than in elementary school. Middle schools include frequent transitions between classes. Students are expected to meet the varying expectations of different teachers in addition to the higher behavioral expectations of this age group. These challenges suggest that developmentally appropriate SWPBS implementation in middle school would be highly beneficial. The current analysis included 13 studies in elementary settings and only three in middle school. This imbalance highlights the priority researchers must give to a redistribution of efforts across a wide range of developmental levels. Generalizability is further limited by the nature of the experimental design.

The school demographic revealed largely consistent findings across subcategories, although the most confidence and largest effect were observed for urban settings. The conclusion can be made that SWPBS has generalizability across different demographics, although – unsurprisingly – it is most effective where the greatest need is for behavioral support. One of the practical features of SWPBS is its flexibility in implementation. Certain features can be scaled up or down given the need of the population. It is unfortunate that few studies reported their implementation in enough detail to categorize this level of intensity. The scaling of treatment potency given student need is another area that lacks thorough investigation.

Finally, it is important to note that one of the lowest effect sizes reported in this meta-analysis was for treatment integrity. Many authors have lamented the lack of attention to treatment integrity.
in the behavioral realm (e.g., Gresham, 1989, 2009; Sheridan, Swanger-Gagne, Welch, Kwon, & Garbacz, 2009). Only 60% of studies included in this synthesis reported a quantitative measure of treatment fidelity (this is a sharp increase, however, from McIntyre et al.’s [2007] finding of 30% for studies published between 1991 and 2005 in the *Journal of Applied Behavior Analysis*). Only one study reported the integrity of the consultative process. Given the vulnerability to low treatment integrity of large-scale system changes in schools, particularly approaches that lack full scripts in implementation, it is important that researchers and practitioners devote significant resources to measurement of this construct. As Gresham (2009) notes, measures of integrity ensure only that the intervention was delivered as intended, not that it was inherently effective. Practically speaking, however, it is likely that modification or erosion of a treatment reduces the potency of the intervention. Without an accurate measure of integrity, the notion of student nonresponsiveness within an RTI model is effectively invalidated by a lack of understanding whether the treatment has been delivered as intended and at the expected dosage.

**Limitations**

Although the Allison-MT method has been shown to have high power and good control for autocorrelation, it tends to inflate effect sizes because of possibly biased error estimates, as can be observed in the distribution of quartile cut-offs reported earlier in this article - far larger than Cohen’s (1988) popular benchmarks. The use of the Allison-MT method in the current study, however, is an important practical inquiry into the nature of single-case effect sizes, which will have a different metric than typical group design studies regardless of the method used. It is likely that more promising alternatives will be generated in the future as the field grows.

This study used both quasi-experimental and experimental single-case designs, limiting the validity that can be derived from our analysis. The use of quasi-experimental designs is sensible in specific cases and will likely continue to be a popular choice for schools. Regardless, further research with better experimental control of individual SWPBS components will greatly add to the body of SWPBS literature.

The use of published dissertations may be considered a weakness of this study. The dissertations, however, did meet our criteria for inclusion, and their use reduces publication bias. Publication bias remains a concern in this study; many large-scale evaluations of SWPBS could not be used in this study due to their reliance on descriptive statistics and lack of baselines. Although program evaluations are useful to implementers of a program, they are limited in their usefulness to potential consumers of SWPBS. Incorporating quasi-experimental methodologies may pose more work for evaluators, but would generate more valid results and would be more generalizable to other consumers.

Finally, two-way interactions could not be calculated for the current study. For example, the effect of SWPBS within specific settings (unstructured vs. structured) may be moderated by the demographic of the school (urban, suburban, rural). The effect of a comprehensive SWPBS program versus the effect of select components may also be moderated by demographic, setting, or age range. These two-way interactions answer important questions regarding where and how SWPBS implementation should occur. Unfortunately, the 20 studies discussed presently are far short of possessing the statistical power required to address this type of moderation. This limited pool of research is further indicative of the need for an expanded research effort on SWPBS.

**Directions for Future Research**

The need for further study on SWPBS cannot be understated considering its widespread implementation. Future efforts need to expand SWPBS effectiveness research beyond elementary schools.
to both younger and older students. If SWPBS implementation precedes this research agenda, it will lack a robust evidence base. Individual components of SWPBS – in isolation and combination – such as active supervision, a token economy, or pre-teaching, need further validation. Eventually the science of school-wide behavioral modification needs to evolve so that we better understand what combination of interventions best matches different levels of student need.

PBS can be delivered at increasing levels of intensity for students with greater behavioral needs when incorporated into a multitiered model. The studies included in this meta-analysis examined PBS in its Tier I form. Many of the studies included did not describe the nature of the behavioral RTI model present, if one existed. This raises the question as to whether PBS at Tier II or Tier III is more effective when SWPBS is present. In addition, it would aid practitioners to know what frequency of non-responders may be expected when SWPBS is implemented.

In summary, the published research on SWPBS has begun to converge into generalizable findings for specific types of DVs and contexts, but there is a pressing need for more research in seldom investigated areas. Avenues for further research include application at the pre-K, middle school, and secondary levels, understanding of the relative cost–benefit of individual components, and greatly extending the duration of intervention phases. Given the popularity of SWPBS, this attention is necessary to qualify SWPBS as evidence-based. Practitioners should take note of the different ways of measuring effectiveness, the rich potential for SWPBS in unstructured settings, and the promise of SWPBS in ameliorating problem behavior and increasing desired behavior in schools.

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis


Blevins, L. D. (2007). The use of school-wide positive behavior support at a rural high school to decrease disruptive behavior for both typical students and students identified with special needs. Dissertation Abstracts International: Section A: Humanities and Social Sciences, 68(12-A), 4924.


Geomatix Ltd. (2005). XYit (3.1.4) [Computer Software].

Psychology in the Schools DOI: 10.1002/pits


*Psychology in the Schools DOI: 10.1002/pits


