Regular Article

Hammill Institute on Disabilities

Remedial and Special Education
2019, Vol. 40(2) 112–128

DOI: 10.1177/0741932517751054
rase.sagepub.com
SAGE

A Systematic Review of Teacher-Delivered Behavior-Specific Praise on K–12 Student Performance

David J. Royer, PhD[1], Kathleen Lynne Lane, PhD, BCBA-D[2], Kristin D. Dunlap, MEd[2], and Robin Parks Ennis, PhD, BCBA-D[3]

Abstract

Behavior-specific praise (BSP) is a core component of many positive behavioral interventions and supports at each level of prevention, often used to increase student academic outcomes and/or reduce inappropriate behavior. We conducted a systematic literature review to explore this low-intensity, teacher-delivered strategy, applying Council for Exceptional Children (CEC) quality indicators and standards to determine whether BSP can be considered an evidence-based practice (EBP). Included articles (N = 6) investigated BSP delivered by a classroom teacher in K–12 traditional school-based settings with academic and/or behavioral student outcome measures. Findings indicated using BSP increased student time on task, decreased inappropriate behaviors, and reduced student tardiness. All studies met our 80% weighted coding criterion. We concluded BSP can be categorized as a potentially EBP based on CEC guidelines. Limitations and directions for future inquiry are presented.

Keywords

behavior-specific praise, Ci3T, low-intensity strategies, PBIS, quality indicators

Praise can have benefits for students and teachers alike, cited as one of the most effective strategies for reducing challenging behavior and promoting prosocial behaviors (Cavanaugh, 2013; Chalk & Bizo, 2004; Sutherland, Wehby, & Copeland, 2000). Benefits for teachers who praised students more often included lower rates of emotional exhaustion and a higher sense of efficacy for classroom behavior management (Reinke, Herman, & Stormont, 2013). Yet, despite the ability of praise to increase teachers' sense of efficacy and reduce/prevent challenging student behavior, teachers reported feeling they do not have the skills to support students with emotional and behavioral needs (Nickerson & Brosof, 2003). This skill gap is an unfortunate barrier to early intervention efforts in general education because strong classroom management and organization are paramount for success within tiered systems of supports (Oliver & Reschly, 2010).

Tiered systems intervene at the first sign of concern, when time and intensity of interventions are lessened through teaching and reinforcing adaptive skills as opposed to reacting after problems occur (Floress & Jenkins, 2015). For example, positive behavioral interventions and supports (PBIS; Horner & Sugai, 2015) and comprehensive, integrated, three-tiered (Ci3T; Lane, Oakes, & Menzies, 2014) models include low-intensity strategies (e.g., behavior-specific praise [BSP], instructional choice, increased opportunities to respond; Lane, Menzies, Ennis, & Oakes, 2015) for primary (Tier 1) prevention, adaptable to Tier 2 interventions for students who need more support (Oakes, Lane, & Germer, 2014). All teachers should be equipped with low-intensity, efficient, easy-to-use strategies to work with students in the classroom when problem behaviors are initially amenable to intervention efforts before prominent academic, behavioral, or social skill gaps appear (Lane & Walker, 2015).

General Praise

Praise in the classroom setting has been studied regularly since White (1975) noted teachers' rate of praise decreased as grade level increased and Brophy (1981) outlined a functional analysis of praise in his seminal work. *Praise* is used instead of *feedback* because it has the added connotation of a more intense response compared with affirming a correct

[1]University of Hawai'i at Mānoa, Honolulu, USA
[2]University of Kansas, Lawrence, USA
[3]University of Alabama at Birmingham, USA

Corresponding Author:
David J. Royer, University of Hawai'i at Mānoa, Wist Hall Rm. 120, 1776 University Ave., Honolulu, HI 96822, USA.
Email: david.royer@hawaii.edu

answer (Brophy, 1981). White reported first and second grades had higher rates of praise compared with other grade levels, the highest observed in second grade at 1.3 approvals per minute. After second grade, the rate of praise dropped sharply and declined until high school, where rates stabilized around one approval for every 5 to 10 min, or four to eight approvals in a 40-min class. In a study spanning first through eighth grades, Brophy described very low rates of praise, less than five praise statements per teacher per hour for academics and less than one praise per teacher per hour for conduct. More recent studies have confirmed similar rates of praise with close agreement across United States, Canada, the United Kingdom, Australia, New Zealand, and Hong Kong (Beaman & Wheldall, 2000; Burnett & Mandel, 2010; Floress & Jenkins, 2015; Reinke et al., 2013). In addition, Brophy argued praise is often used incorrectly and should not be assumed to be reinforcing as teachers most often do not praise strategically or with intent to reinforce student behavior. To praise correctly, Brophy based recommendations on O'Leary and O'Leary (1977): Praise should be contingent on student behavior, specific, sincere, varied, and credible. Yet, praise has been found to rarely be contingent, rarely used as positive reinforcement, and rarely observed with student behavior (Beaman & Wheldall, 2000).

BSP

BSP is one strategy teachers can use daily to prevent and reduce challenging behaviors. Unlike general praise, with BSP, educators say or write the precise behavior exhibited and how it met an expectation or affected academic/social achievement (Kennedy & Jolivette, 2008). The student must be explicitly told what malleable factor within the student's locus of control is being praised (e.g., "Good job studying for this science test, your effort paid off"), rather than uncontrollable factors such as intelligence (e.g., "You're so smart") or ability (e.g., "You're a natural-born leader"). Specifically, praising effort instead of ability may help students attend to the method of tasks and be motivated by the opportunities and potential hard work may bring (Weaver & Watson, 2004). BSP should be sincere so if the student finds attention reinforcing the praised behavior is likely to reoccur (Lane, Menzies, et al., 2015).

Teachers within tiered systems of supports can use BSP to recognize students who meet school-wide expectations posted in all key settings, reinforcing desired behaviors while also reminding struggling students of current expectations. Recognizing expectations met with BSP helps orient all students to the behaviors teachers expect students to engage in to be successful (Brophy, 1981). Students whose behavior consistently does not meet defined expectations can receive higher rates of BSP as a Tier 2 intervention, and BSP can be a component in any Tier 2 or 3 intervention. For example, Caldarella, Christensen, Young, and Densley (2011) implemented a praise note system that quickly reduced variability within and number of student tardies. Consistent delivery of a praise note each time a student was on time was not required for results to manifest quickly. As illustrated, appropriate behavior should be supported by BSP before more intensive interventions are considered (Stormont & Reinke, 2009).

Natural Rates of BSP

Researchers have recently observed general and BSP in classrooms, reporting higher BSP rates than the 5% reported by Brophy (1981) but still significantly lower than general praise. Among four kindergarten teachers, Floress and Jenkins (2015) recorded 38.5 general praises per hour and 1.75 BSP per hour. For Grades K–3, Reinke et al. (2013) found 25.8 general and 7.8 BSP per hour. In Australia, Burnett and Mandel (2010) observed Grades 1 to 6 and reported 29 general praises per hour and 1.75 BSP per hour.

Rates of praise in special education classrooms or for students with disabilities in general education classrooms were even lower. Gable, Hendrickson, Young, Shores, and Stowitschek (1982) collected data for 97 teachers of students with learning disabilities, emotional and behavioral disorders (EBD), intellectual disability, and multiple disabilities, and found general praise occurred 4.4 to 13.5 times per hour. Shores et al. (1993) reported a rate of 4.49 praises per hour for aggressive students with EBD in self-contained special education classrooms and 0.42 per hour for nonaggressive students in general education classrooms. Wehby, Symons, and Shores (1995) similarly found students with EBD in self-contained special education classes received 1.35 (low aggressors) or 2.35 (high aggressors) praises per hour. No studies investigated naturally occurring (nonexperimental) rates of BSP in special education settings (Jenkins, Floress, & Reinke, 2015). Clearly, there is a need for special education teachers to use higher rates of praise, and to be an effective reinforcer, such praise should be behavior specific.

Benefits of BSP

As illustrated, teachers naturally use praise to varying degrees, most often as general praise. Praise can be more effective when it is specific and used intentionally as positive reinforcement (Brophy, 1981; Thompson, Marchant, Anderson, Prater, & Gibb, 2012). General praise rarely translates to improved on-task behavior, assignment understanding, or self-confidence (Hattie & Timperley, 2007). In contrast, BSP provides students specific performance feedback (academic, behavioral, or social), helps students realize what they have specifically done well, reinforces school-wide expectations, and can make the future occurrence of

socially acceptable behavior more likely, following tenets of applied behavior analysis (ABA; Brophy, 1981; Cooper, Heron, & Heward, 2007; Lane, Menzies, et al., 2015; Sutherland et al., 2000).

Establishing an Evidence Base for BSP

Despite widespread use of praise and studies exploring the functional relation between BSP and disruptive behavior since Madsen, Becker, and Thomas (1968), there are few reviews exploring the body of evidence behind BSP. In a literature review conducted by Simonsen, Fairbanks, Briesch, Myers, and Sugai (2008), 20 general classroom management practices meeting criteria (i.e., methodologically sound experimental design, demonstrated effectiveness, supported by three or more studies) were identified, including specific and/or contingent praise. Cavanaugh (2013) reviewed the use of performance feedback for teachers' use of BSP and opportunities to respond. Cavanaugh identified 24 articles with strategies to increase teacher use of BSP, including performance feedback, training, modeling, goal setting, self-monitoring, and role-play. Although most studies demonstrated these strategies increased teacher use of BSP, Cavanaugh did not explore effects of increased praise on student-level outcomes, focusing instead on teacher-level outcomes.

Purpose

Previous reviews established a foundation for understanding the body of BSP literature. Both reviews were published prior to Council for Exceptional Children's (CEC) *Standards for Evidence-Based Practices in Special Education* (CEC, 2014), guidelines for determining methodologically sound research and the evidence base for a practice. We were specifically interested in determining whether BSP is an evidence-based practice (EBP) according to CEC (2014) guidelines. This systematic literature review examined the evidence base for teacher-delivered BSP as an intervention for increasing academic, behavioral, and social success and/or for reducing problem behaviors, limiting our search to traditional K–12 educational settings. We focused on traditional educational settings, including typically developing school-age students and those at risk of, or with, identified disabilities, reflecting today's diverse classroom. We believe the classroom is where evidence is needed to show which low-intensity interventions can be effective in daily use by practitioners to increase academic achievement and reduce challenging behavior. Our research questions were as follows:

Research Question 1: To what extent did teacher-delivered BSP intervention studies address CEC (2014) quality indicators (QIs)?

Research Question 2: What is the nature of the evidence base supporting BSP according to CEC (2014) guidelines, applying an 80% minimum criterion for methodologically sound studies (Lane, Kalberg, & Shepcaro, 2009)?

Research Question 3: What was the magnitude of effects for teacher-delivered BSP interventions?

Method

Search and Article Selection

We searched 21 electronic databases to identify BSP intervention studies with pre-K–12 students. Databases were ABI/INFORM Global, Academic Search Complete, Educational Resources Information Center (ERIC), JSTOR Archival Journals, Linguistics & Language Behavior Abstracts, MEDLINE/PubMed, MLA International Bibliography, OneFile, Project MUSE, ProQuest Nursing & Allied Health Source, ProQuest Research Library, PsycARTICLES, PsychINFO, SAGE Journals, Science Citation Index Expanded (Web of Science), SciVerse ScienceDirect (Elsevier), Social Sciences Citation Index (Web of Science), Sociological Abstracts, SpringerLink, Taylor & Francis, and Wiley Online Library. Boolean search terms were used to include combinations and derivatives of *behavio** AND *specific* AND *praise*, "*positive verbal praise*," and (*teacher* OR *peer*) AND "*praise notes*." This search returned 318 results, replicated with 100% accuracy by a second author for reliability (see Figure 1).

Using a binary coding system where 0 = *does not meet inclusion criteria* and 1 = *meets inclusion criteria*, two authors independently read the 318 article titles and abstracts to determine which articles should be read in full to see whether inclusion criteria (described subsequently) were met. Fifty-five articles from 39 journals were selected for reading in full. Interrater agreement (IRA) between readers was 87.11% ($\kappa = .56$, 95% confidence interval [CI] = [0.44, 0.69], indicating moderate agreement; Landis & Koch, 1977), calculated by dividing the number of point-by-point agreements by 318 and multiplying by 100, with disagreements discussed and resolved. Next, we independently read the 55 articles in full and retained 33 from 22 journals (IRA = 85.45%; $\kappa = .69$, 95% CI = [0.49, 0.89], indicating substantial agreement).

Independent ancestral searches of the 33 articles' references revealed 32 titles of interest (IRA = 96.13%; $\kappa = .81$, 95% CI = [0.76, 0.87], indicating near perfect agreement). After reading abstracts, 13 were selected to read in full (IRA = 75.00%; $\kappa = .46$, 95% CI = [0.13, 0.78], indicating moderate agreement), and six met inclusion criteria (IRA = 92.31%; $\kappa = .85$, 95% CI = [0.56, 1.00], indicating near perfect agreement), bringing the total to 39 articles from 25 journals.

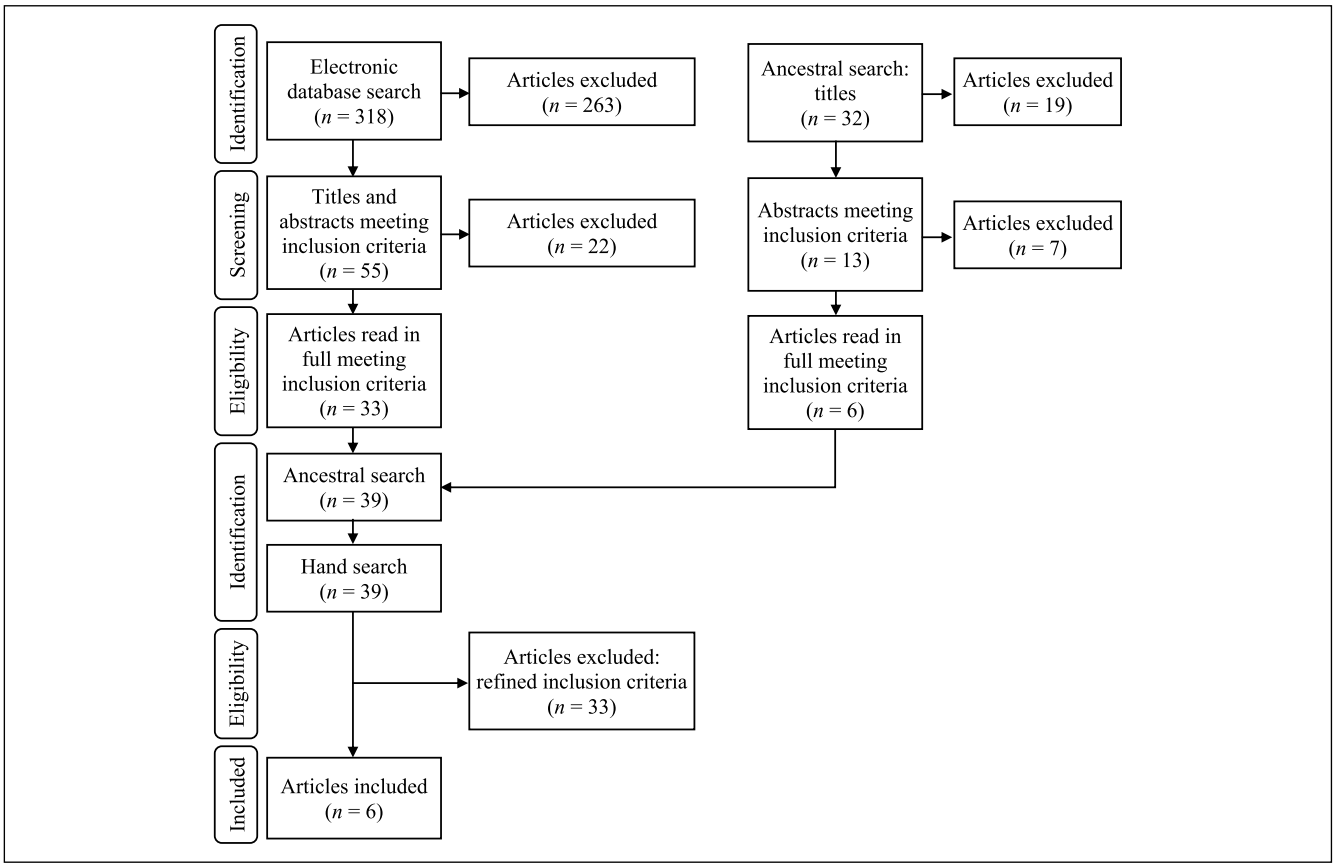

Figure 1. Flow diagram illustrating search procedures and article inclusion for behavior-specific praise systematic literature review.

We then conducted independent hand searches of the local university library stacks. January 1968 to March 2016 journals with two or more of the 39 included studies were hand searched to verify electronic search results and identify additional articles. Journals identified for hand searching were *Behavior Modification, Behavioral Disorders, Beyond Behavior, Education & Treatment of Children, Journal of Applied Behavior Analysis, Journal of Behavioral Education, Journal of Positive Behavior Intervention*, and *Preventing School Failure* for a total of 965 issues. Volumes not available in the library stacks were searched online by clicking through each issue in similar manner to bound volumes. No additional articles were identified in this step. Mean IRA across all hand-searched journals was 98.55% (range = 95.77%–100%; κ = .77, 95% CI = [0.62, 0.93], indicating substantial agreement).

After hand searching, we reviewed the 39 articles and determined the independent variable (IV) and participants across studies were heterogeneous enough to warrant (a) exclusion of articles where BSP was part of a packaged intervention (n = 12) such as precorrection + BSP (e.g., Smith, Lewis, & Stormont, 2011), instead focusing our review on articles where BSP was the main IV; (b) exclusion of articles where experimental design (e.g., Chalk & Bizo, 2004) or method and results sections (e.g., Caldarella et al., 2011) were not clearly defined; and (c) inclusion of only studies where teacher-delivered BSP was the IV (e.g., excluded coaching of teachers to increase BSP, student-delivered BSP). We continue this review with six articles meeting inclusion criteria (IRA = 100%) for teacher-delivered BSP as the IV.

Finally, we contacted corresponding authors and journal editors of included studies and asked whether they had any studies accepted for peer review, in press, or published recently not found by our searches. No additional articles met inclusion criteria for teacher-delivered BSP.

Inclusion Criteria

Included studies met five criteria: (a) IV was teacher-delivered BSP, defined as "providing students with praise statements that explicitly describe the behavior being praised" (Allday et al., 2012, p. 87). We excluded studies if the intervention agent was only described as the experimenter and was not clearly the classroom teacher (e.g., Weaver & Watson, 2004) or if sessions were conducted outside the classroom in a clinical format (e.g., Polick, Carr, & Hanney, 2012). (b) Student dependent variables (DVs) included

variations of disruptive and problem behavior, stereotypy, time on task/academic engaged time, work completion, work accuracy, and/or social skills. (c) Interventions occurred with K–12 students in traditional school-based settings. Discrete trial training and studies in residential treatment centers (e.g., Kennedy & Jolivette, 2008), home settings, or clinics resembling classroom settings (e.g., Everett, Olmi, Edwards, & Tingstrom, 2005) were not considered due to the highly controlled nature of such settings and/or substantial variability from traditional school settings. (d) The study followed an experimental design. (e) The article was in English and published in a refereed journal. Although the peer review process is not perfect protection against errors, it is one essential element of scientific inquiry, increasing probability of accurate analysis (Resnik & Elmore, 2016).

Coding Procedures

Training. The first and third authors were trained in QI coding by the second author. Training consisted of reading multiple sources for QIs (i.e., CEC, 2014; Gersten et al., 2005; Horner et al., 2005; What Works Clearinghouse, 2013) and coding practice articles using a QI matrix (Lane, Common, Royer, & Muller, 2014). Practice article coding was compared and disagreements discussed and resolved, with criterion set at three consecutive articles ≥85% IRA. For single-case research designs (SCRDs), the first author's mean IRA with the second author was 96.67% (range = 90.00%–100%).

Descriptive coding. We coded descriptive characteristics of each article into Table 1, including (a) context and setting, (b) participants, (c) intervention agent, (d) description of practice, (e) implementation fidelity, (f) internal validity, (g) outcome measures/DVs, and (h) data analysis (IRA = 95.18%).

QI coding. Two authors read and coded included articles for the presence or absence of QIs of methodologically sound interventions as defined by CEC (2014). The eight QIs coded were (1.0) context and setting, (2.0) participants, (3.0) intervention agent, (4.0) description of practice, (5.0) implementation fidelity, (6.0) internal validity, (7.0) outcome measures/DVs, and (8.0) data analysis. Authors compared independent coding, resolved discrepancies (n = 12 of 168), and calculated IRA using point-by-point correspondence (Cooper et al., 2007) for each article and for each QI component across studies by dividing the sum of agreements by the number of components (i.e., 22 for SCRDs) or by six (number of articles), respectively, multiplied by 100 to obtain a percentage. Mean IRA for articles and QI components was 90.91% (article range = 86.36%–95.45%; component range = 50.00%–100%; κ = .83, 95% CI = [0.73, 0.92]).

Methodological QIs

QI 1.0. Context and setting. To meet QI 1.1, a study had to describe critical features of context/setting (e.g., public school, university laboratory school, general education classroom, self-contained classroom for students with EBD) in sufficient detail to allow reviewers to determine whether it should be included in their review. IRA for QI 1.1 was 100%.

QI 2.0. Participants. To meet QI 2.1, a study had to describe relevant participant demographics to identify a population to which results might be generalized. We considered this met if at least one demographic element was reported for each participant or a range when a whole class was the unit of analysis. To meet QI 2.2, a study had to describe participant disability or risk status and determination method (i.e., who applied what criteria). When individual students were participants, stating the teacher nominated them based on academic or behavior concerns was not sufficient unless additional detail enabling replication was provided (e.g., quantitative information from systematic screening, specific behaviors confirmed by secondary observers). If the unit of analysis was the whole class, we required the study to state the type of classroom. IRA for QI 2.1 was 83.33% and 2.2 was 50.00% (see the "Discussion" section).

QI 3.0. Intervention agent. To meet QI 3.1, authors described intervention agent role and preferably background variables, though the latter were not relevant to our research questions. QI 3.2 required (a) description of specific training or qualifications when required and (b) indication the intervention agent met training criterion or possessed required qualifications to implement the intervention. IRA for QI 3.1 was 100% and 50.00% for QI 3.2.

QI 4.0. Description of practice. Authors of included studies had to describe intervention procedures and intervention agent actions as well as study materials (or cite) with sufficient detail to allow replication by other researchers to meet QI 4.1 and 4.2, respectively. A general description of materials, such as "list of praise statements," also needed to provide examples to meet QI 4.2 for our review. IRA was 100% for QIs 4.1 and 4.2.

QI 5.0. Implementation fidelity. Fidelity related to adherence (QI 5.1) was met when authors used a direct, reliable measure such as an observation checklist of intervention components. Included articles' direct observation of teachers' use of BSP was sufficient. Implementation fidelity related to dosage (QI 5.2) was also met by use of a direct, reliable measure. We coded QI 5.2 met if dosage was described (e.g., 30-min daily sessions over 4 weeks) along with a graph of DV data showing total number of sessions to allow

Table 1. Descriptive Results and Effect Classification of Included BSP Studies.

CEC (2014) Quality indicator	Madsen, Becker, and Thomas (1968)	van der Mars (1989)	Houghton, Wheldall, Jukes, and Sharpe (1990)	Wheatley et al. (2009)	Haydon and Musti-Rao (2011)	Hollingshead, Kroeger, Altus, and Trytten (2016)
1.0. Context and setting	Public elementary; Classroom A = second grade, 29 students; Classroom B = K, 20 students	Second grade PE class with 14 students held indoor twice a week for 30 min	Large inner-ring comprehensive school in the West Midlands, England	Rural northern Utah ES, 200 students Grades 1–5, 80% White, predominant mid to lower-mid SES; lunchroom intervention setting, three grades at a time	Large Midwestern city; two math classes, public MS (Grades 7–8), 700 students, 55% White, 42% Black, 92% FRL, 11% transient, no AYP 2006–2009	Inner-city urban HS with ~750 students: Black (86%), White (9.4%), multiracial (3.4%), FRL (84%), SpEd (26%); seventh-grade social studies class
2.0. Participants	Three students: Cliff, second grade, IQ 123; Frank, second grade, IQ 106 on Stanford-Binet; Stan, K; selection = volunteer teachers' classrooms observed two students with high frequency PB selected from each (one moved after baseline)	Two boys, one girl in second grade; identified by teacher as frequently off task (inattentive, inappropriate use of equipment, talkouts, peer interaction during instructions, not following directions, not completing transition tasks)	Four classrooms of above avg. ability students: A = third-year English, 13 boys, 14 girls; B = second-year geography, 12 boys, 13 girls; C = fifth-year math, 12 boys, 14 girls; D = fourth-year chemistry, 13 boys, nine girls; selection criteria = teacher volunteers	School-wide intervention: 200 students Grades 1–5, 80% White, predominant middle to lower middle SES; selection criteria = part of school-wide effort to decrease PB in the lunchroom	20–22 students each class; demographics not described; classes selected by principal based on teacher request for CM strategies	25–31 seventh-grade social studies students, 17 male, 14 female, primarily Black, 1 White, 6 with IEPs (5 SLD, 1 ED), 3 at risk due to number of office discipline referrals; school selected for not having PBIS, to examine effectiveness of PBIS-based interventions
3.0. Intervention agent	Two teachers, Mrs. A and Mrs. B; training = individual conferences to explain provided written instructions for each phase, plus weekly seminars on behavior analysis and specific suggestions for situations based on observations	Male PE teacher, 5 years' experience; training not described	Four secondary teachers: three male, one female, 13 years avg. experience; training = videos on effective TR and praise (give briefly, private, eye contact, specific), given handout, instructed to reduce TR and increase BSP to 3.2 or 4.2 per 15 min	Teachers, custodians, lunchroom staff, school admin; overview training on praise notes; 20 min training with students before each phase, in lunchroom with modeling, role-play, and practice, student checks for understanding, and reinforcement	Two first-year math teachers, grad degree or certification in math, male, White, 33–34 years of age, 1 had undergraduate CM course; individual 40-min training to use BSP with script, example BSP statements, MotivAider set at 4 min, and 16-min practice session	Seventh-grade White female teacher with 17 years of experience across six different schools and a master's degree; training = 30-min session on using BSP, designing classroom rules, reducing TR, ignoring negative behavior
4.0. Description of practice	Rules phase = posted 5–6 rules, reviewed avg. 5.2 times per day; ignoring phase = ignore behavior unless hurting others; praise phase = give BSP, attention, or smile when expectations met; Class A phases were additive, returned to baseline, then back to rules + ignoring + praise combo; Class B rules and ignoring phases were distinct but final phase combined all three	Teacher increased BSP for target students' PE class conduct and skill performance when hearing audio reminders via earphone at an avg. rate of 1.7 per min (range = 0.51–1.23 per min)	Teachers reduced the number of and how TR were given (brief, explicit, private) and increased the number of and how praise was provided (brief, specific, private) in different conditions, using a hand tally counter to self-report	Three phases targeting littering, inappropriate sitting, and running in the lunchroom; faculty and staff gave verbal BSP every 2–3 min for appropriate behavior targeted in the current phase and gave written praise note, which contained student and teacher signature lines, mascot, and motto: "safe, kind, and responsible"	Place script by overhead projector, when MotivAider vibrates every 4 min deliver BSP for appropriate behavior	Teacher trained on using BSP, designing rules, reducing TR, and ignoring negative behavior; worked with students to establish five to six positively stated rules displayed in class; teacher used BSP for behavior already displayed in class with competence
5.0. Implementation fidelity	Teacher verbal behaviors to all children (not just to target children) rated on avg. once per week with 50% assessed by second rater (IRR = 85%, range = 70%–96%)	Teacher wore wireless microphone to record verbal behavior and PE classes were videotaped for data collection; event recording used videotape running stopwatch timer; mean rate of teacher praise per subject per phase reported; IOA range for teacher behavior = 92.3%–100%	Teachers self-recorded praise and TR counts; researchers used Observing Pupils and Teachers in Classrooms observation tool (section A = teacher positive, negative behavior; section B = student on-task behavior); mean IOA for four teachers' behavior = 94%, 99%, 92%, 100%	Primary lunchroom observer randomly chose two staff members on 61% of intervention days to check at least one praise note was delivered during a 3-min period, training procedure was followed, and praise note was completed correctly; staff met expectations 100% of observations	25-min direct observation of TR, GP, and BSP for 100% of sessions (2 days per week over 8 weeks); both teachers implemented BSP every 4 min with 100% fidelity; mean IOA for teacher BSP = 89.5% (range = 80.0%–100%), TR = 86.0% (range = 81.3%–100%)	BSP and TR data collected each observation session and teacher was provided daily graphed data, comments, and field notes; IOA for teacher behavior = 100%

(continued)

Table 1. (continued)

CEC (2014) Quality indicator	Madsen, Becker, and Thomas (1968)	van der Mars (1989)	Houghton, Wheldall, Jukes, and Sharpe (1990)	Wheatley et al. (2009)	Haydon and Musti-Rao (2011)	Hollingshead, Kroeger, Altus, and Trytten (2016)
6.0. Internal validity	Classroom A = A-B-BC-BCD-A-BCD withdrawal design; Classroom B = A-B-C-BCD design; during baseline teachers made not changes to behavior, class was run their typical way	MBD across three participants; baseline rate of praise = 0.40, 0.36, and 0.44 for Students 1, 2, and 3, respectively	Two MBD sets with two teachers per set, with follow up for 3/4; baseline = variable/declining on-task behavior, Teachers A & B = higher TR rates than praise, C & D = low rates of TR, higher praise rates	MBD across three target behaviors; during baseline no program targeted lunchroom behavior, staff circulated and corrected student behavior at will	Multiple-probe baseline design; during baseline teachers asked questions and occasionally volunteers worked problems on white board or with calculators	A-B-A-B design with maintenance phase; during baseline teacher BSP (0) and TR (1.5+ per min) statements were recorded, plus student on- and off-task behavior
7.0. Outcome measures/DVs	DV = PB: out of seat, moving chair, object noise, disturbance of other's property, physical contact, verbalization during nonpermitted times, turning around, mouthing objects, isolated play (K only), ignore teacher, playing with objects, engaging in something different than directed; SV not measured	DV = off-task behavior measured with Academic Learning Time-PE version 2 (Siedentop, Tousignant, & Parker, 1982) including behavior disruptions, misbehavior, talking during instructions, fighting, and disrupting a drill (mean IOA across students = 88.2%–94.0%); IOA for BSP = 92.3%–100%; SV not measured	DV = on-task behavior defined by citing Merrett and Wheldall (1986): getting on with work set by the teacher and/or following his or her instructions (mean IOA across teachers = 92%–100%); SV not measured but demonstrated through continued use of intervention by teachers at follow up observations	DVs event recording = littering: objects larger than 3 × 5 index card (IOA for 43% of sessions = 98%); inappropriate sitting: standing or back pockets not on seat (IOA for 28% of sessions = 90%); running: having both feet leave floor at the same time, including hopping, skipping, jumping (IOA for 28% of sessions = 87%); SV not measured	DVs = DB event recording within 10-s intervals, mean IOA for 25%–33% of sessions = 97.6% (range = 95.6%–98.7%); SV = teacher interview: BSP not difficult, MotivAider was not intrusive, they might be very likely to use BSP in the future, and training sessions were very helpful	DV = on-task behavior defined as staying in one's seat and completing work, recorded by whole interval, rotating between tables of students every 15 s (IOA for 31% of sessions ranged from 88–100% across all phases); SV measured through anecdotal comments and strong desire to return to BSP during withdrawal phase
8.0. Data analysis	Praise increased 16.5% for both teachers; major changes to PB occurred only when praise was in effect (t test $p < .05$, $df = 2$); rules alone had little effect; ignoring alone or with rules had inconsistent results; avg. of two students' % of intervals with PB went from 46.8% to 20.5% baseline I to praise I, 37.6% to 15.1% baseline II to praise II; no functional relation established: no replications after demonstrations	Praise increased from baseline rate of 0.40, 0.36, and 0.44 to 1.06 (265%), 0.90 (250%), and 0.69 (156%) during intervention for Students 1, 2, and 3, respectively; student off-task behavior ranged from 26.5%–39.4% of intervals during baseline and dropped to 9.3%–15.4% during intervention	Praise increased from baseline avg. for all teachers of 4.5 per 15 minute to avg. of 7.1 per 15 minute during phases with BSP; TR decreased from baseline avg. for all teachers of 8.1 to 2.8; student on-task behavior in the four classes increased an avg. 22% (range = 19%–27%) baseline to intervention II; functional relation not established: different order of intervention phases in MBD for each set of teachers	Litter decreased from avg. of 34.3 pieces to 1.3 (96% reduction); inappropriate sitting decreased from avg. of 65.5 to 23.3 (65% reduction); running decreased from avg. of 34 instances to 8.5 (75% reduction)	Mr. Smith: BSP increased from 0 to 0.3/min, GP increased from 0 to 0.2/min, TR decreased from 1.28/min to 0.25/min, DB decreased from 2.1/min to 0.49/min; Mr. Jones: BSP increased from 0 to 0.7/min, GP increased from 0 to 0.3/min, TR decreased from 0.86/min to 0.21/min, DB decreased from 1.6/min to 0.7/min; functional relation not established due to only two cases within multiple-probe baseline design	BSP increased from 0 occurrences in baseline and withdrawal to a mean of 29.2 (range = 10–46) during intervention I and 18.6 (range = 11–26) during intervention II; student on-task behavior increased from a baseline mean of 52.6% to 74.2% in intervention I, dropped to 57% in withdrawal, increased to 70.8% in intervention II, and maintained at 67.5%

Note. CEC = Council for Exceptional Children; K = kindergarten; PE = physical education; ES = elementary school; SES = socioeconomic status; MS = middle school; FRL = free or reduced-price lunch; SpEd = special education; AYP = adequate yearly progress; HS = high school; IQ = intelligence quotient; PB = problem behavior; avg. = average; CM = classroom management; IEP = individualized education program; SLD = specific learning disability; ED = emotional disturbance; PBIS = positive behavioral interventions and supports; exp. = experience; TR = teacher reprimands; BSP = behavior-specific praise; IRR = interrater reliability; IOA = interobserver agreement; GP = general praise; MBD = multiple baseline design; DB = disruptive behavior; DV = dependent variable; SV = social validity.

calculation. To meet QI 5.3, authors assessed fidelity throughout the intervention (e.g., beginning, middle, end) for each unit of analysis. We did not require fidelity measurements for each condition/phase if an aggregate from different time points was provided. IRA was 100% for QIs 5.1, 5.2, and 5.3.

QI 6.0. Internal validity. The internal validity QI consists of nine components, with six applicable to SCRD (no group studies met inclusion criteria). First, for QI 6.1, investigators had to demonstrate control and systematic manipulation of the IV. We determined this QI could only be met if implementation fidelity (QI 5.1) was met, allowing the possibility of experimental control (Slaughter, Hill, & Snelgrove-Clarke, 2015). In other words, it was not possible to verify the IV was under the control of the researcher if there was no method to determine the intervention was implemented as designed. Second, a description of baseline was required for QI 6.2. Because BSP is a strategy applicable to all content areas and interventions, we required articles only describe baseline rate of BSP and DVs and at least one general student-level item (e.g., during math). Third, participants had to have no or very limited access to intervention components during baseline and withdrawal phases to meet QI 6.3. We required authors state this explicitly (e.g., teachers received no training on BSP prior to the first intervention phase).

A design that allowed for the possibility of three demonstrations of experimental effect at three different times (e.g., A-B-A-B, multiple baseline) met QI 6.5 and we confirmed graphed results matched stated design. To meet QI 6.6, studies with a baseline phase included at least three baseline data points. Exceptions could be made if justification was provided, such as ethical considerations (e.g., behavior was harmful) or a counter-therapeutic trend, but it was not sufficient to state withdrawal had less than three data points for reasons of convenience. QI 6.7 required control for common threats to internal validity, accomplished through use of SCRD. IRA was 100% for QIs 6.1, 6.3, 6.6, and 6.7, 66.67% for QI 6.2, and 83.33% for QI 6.5.

QI 7.0. Outcome measures/DVs. QI 7.0 has six components, with 7.6 applicable only to group designs. For QI 7.1, outcomes of the intervention had to be socially important. We accepted a social validity questionnaire or a strong argument in the study's introduction or discussion. For 7.2, a study had to define DVs and clearly describe their measurement. For 7.3, all outcome measures had to be reported (clearly graphed data were sufficient). For 7.4, frequency and timing of outcome measures had to be appropriate. A minimum of three data points per phase was required (or otherwise justified). For 7.5, interobserver agreement (IOA) had to be ≥80% or kappa ≥60%. We considered mean IOA acceptable if range was reported and was not less than 60%. IRA for all 7.0 QIs was 100% except QI 7.4, which was 66.67%.

QI 8.0. Data analysis. SCRDs had to include graphs of student outcome data clear enough for visual analysis to determine experimental control to meet QI 8.2. QIs 8.1 and 8.3 were only applicable to group comparison studies. IRA for QI 8.2 was 100%.

Evaluation Procedures for Classifying the Evidence Base of Practices

After coding, we followed CEC (2014) procedures to ascertain whether BSP met criteria for an *EBP, potentially EBP, mixed effects, insufficient evidence*, or *negative effects*. To be eligible to contribute to the evidence base, an SCRD study must be methodologically sound (which we defined as meeting 80% of QIs; Lane et al., 2009), have a minimum of three cases with 75% demonstrating therapeutic change as a result of a functional relation between the IV and DVs, and have positive or neutral/mixed effects based on the direction of functional relationship. Determination of the EBP category is then based on the number and type of methodologically sound studies and their effects (see CEC, 2014, for details).

We used a weighted coding method to recognize high-quality articles that met 80% or more QI components by proportionally weighting scores to contribute to a composite score (Lane et al., 2009). For example, QI 7.0 has five components applicable to SCRD, so each component contributes 20% to the total score for QI 7.0. If four of the five components of QI 7.0 were met, rather than score the QI as zero for not met, weighted scores for remaining components contributed to a composite score of 0.80 (0.20 + 0.20 + 0.20 + 0.20 = 0.80). A total score of 6.40 (80% × 8 QIs = 6.40) identified articles as methodologically sound.

Data Extraction and Analysis

We extracted data from graphs using WebPlotDigitizer (Version 3.11; Rohatgi, 2017) and performed three further analyses. First, the percentage of nonoverlapping data (PND) between phases was calculated (Scruggs & Mastropieri, 2013). To calculate an overall PND for a study with A-B-A-B reversal designs and/or multiple cases, we combined data overlap across phases and cases (i.e., total number of nonoverlapping data points from all participants' intervention phases divided by total number of intervention data points). This aggregated the proportion of data in each intervention phase that showed improvement beyond the highest/lowest (depending on intended effect) datum point in the preceding baseline phase. Higher PND indicated a larger intervention impact, with <50% considered ineffective, 50% to 70% questionable, 70% to 90% effective, and

≥90% considered very effective (Scruggs & Mastropieri, 1998).

Second, we calculated Tau-U using an online calculator (Vannest, Parker, Gonen, & Adiguzel, 2016), producing a weighted average Tau-U across participants or cases for each study. Tau-U is a conservative nonparametric statistic appropriate for small *n* SCRDs that controls for positive baseline trend (which we defined as 20% or greater of all baseline data pairs showing improvement over time, i.e., monotonic trend), includes phase contrast and intervention trend in the calculation, and has strong statistical power (Parker, Vannest, Davis, & Sauber, 2011; Vannest & Ninci, 2015). Tau-U has a variable range, determined by the length of baseline and intervention phases (Pustejovsky, 2016a, 2016b). According to Vannest and Ninci (2015), Tau-U can be interpreted as 0 to .20 = small, .20 to .60 = moderate, .60 to .80 = large, and greater than .80 very large, depending on context, social validity, and needs of the participants.

Third, we calculated between-case standardized mean difference (BC-SMD; Hedges, Pustejovsky, & Shadish, 2012, 2013) effect sizes for the five included studies where the three participant minimum criterion was met (Valentine, Tanner-Smith, Pustejovsky, & Lau, 2016). BC-SMD effect sizes are comparable with standardized mean differences from between-group experimental designs (e.g., Cohen's *d*). Single-case data are modeled with a hierarchical linear approach to take into account the nested structure of SCRD (Valentine et al., 2016). BC-SMD effect sizes can be interpreted following other standardized mean difference effect size interpretation (e.g., Fritz, Morris, & Richler, 2012): small (0.20–0.50), medium (0.50–0.80), large (≥0.80).

Prior to extracting data, the primary coder coached the secondary coder (approximately 30 min) through each step of a task analysis (see Drevon, Fursa, & Malcolm, 2016) of data extraction using a graph from a study not included in this review. When complete, coders independently practiced data extraction using three additional graphs (each with 19 data points) from a study where actual values were known. Extracted data were compared with known data values, differences computed point by point, and the mean difference for each coder for each graph calculated. Overall means for each coder were compared by *t* test and no significant difference was found, $t(3) = 0.21$, $p = .85$. In addition, each extracted datum point was compared with the corresponding known value and was found to be within ±5% (except for zero, where 5% could not be calculated) and no extracted value was off by more than 0.75 (range = 0.00–0.74).

Results

Descriptive Characteristics

The six included studies ranged in publication from 1968 to 2016, each from a unique journal. One study (Houghton, Wheldall, Jukes, & Sharpe, 1990) was conducted in England, the rest in the United States. Studies were conducted in four elementary schools, one middle school (Haydon & Musti-Rao, 2011), and one Grades 7 to 12 high school (Hollingshead, Kroeger, Altus, & Trytten, 2016), with BSP implemented for the entire class or school-wide (Wheatley et al., 2009). Two studies (Madsen et al., 1968; van der Mars, 1989) collected data for target students instead of the whole class, though BSP was delivered class wide (see Table 1).

Methodological QIs

1.0. Context and setting. All studies met QI 1.0, describing context and setting. Authors reported school and classroom type (e.g., elementary physical education [PE]) but not always if the school was public or private. Three studies provided school demographics including total students, ethnicity, and socioeconomic status (Haydon & Musti-Rao, 2011; Hollingshead et al., 2016; Wheatley et al., 2009); two reported geographic location (Haydon & Musti-Rao, 2011; Wheatley et al., 2009); Hollingshead et al. (2016) reported the school's percentage of students receiving special education services; Haydon and Musti-Rao (2011) reported transience rate.

2.0. Participants. All studies met QI 2.1 participant demographics, and five (83.33%) met QI 2.2 risk status. Although we did not require studies where the classroom was the unit of analysis to describe how each class was selected for participation, Haydon and Musti-Rao (2011) described how each classroom was observed after teachers volunteered to confirm classes had high frequency problem behavior. Similarly, Madsen et al. (1968) observed student behaviors after teachers volunteered to participate in the study before selecting them for inclusion.

3.0. Intervention agent. All studies met QI 3.1 by describing the interventionist's role and though not relevant to the focus of our review, with one exception also providing background variables. Although five studies described training procedures, none met QI 3.2 because none reported a check for understanding or a training criterion was met.

4.0. Description of a practice. All studies met both QIs 4.1 and 4.2, sufficiently describing intervention procedures and materials. For example, van der Mars (1989) described how the PE teacher delivered BSP class wide when prompted by auditory reminders delivered at a mean rate of 1.7 per minute. Audio cues were heard via a mini-earphone attached to a microcassette player worn in a small pouch at the waist, which also contained a wireless microphone for recording teacher verbal behavior, and the whole class was video recorded for IV and DV data collection. Wheatley et al.

(2009) provided rich description of written praise notes faculty gave students (i.e., small, yellow, signature lines, mascot, motto), and how faculty participation was increased by entering teacher names from the weekly winning student praise notes into drawings. All praise notes were stapled to the entrance hallway board weekly, and when full, earned a larger school-wide reward.

5.0. Implementation fidelity. All studies met all QI 5.0 components for reporting implementation fidelity. Five focused on verbal BSP (compared with written praise) as the IV and used direct observation (in vivo or video recording) to regularly track treatment integrity, with all reporting IOA data. van der Mars (1989) reviewed video with event recording using the on-screen timer to calculate rate of teacher BSP per subject per phase. An independent observer reanalyzed two class recordings from each phase and calculated agreement percentages.

6.0. Internal validity. All studies met QIs 6.1, 6.2, 6.3, and 6.6. Three (50.00%) met QI 6.5 and four (66.67%) met QI 6.7, whereas QIs 6.4, 6.8, and 6.9 were not applicable to SCRDs. In addition to describing baseline rate of teacher BSP for QI 6.2, Haydon and Musti-Rao (2011) also reported previous interventions attempted to increase effective teaching strategies and student outcomes, and described typical teacher and student routines and interactions. Wheatley et al. (2009) met QI 6.3 (baseline isolated from intervention) by stating during baseline no school programs attempted to improve student behavior in the lunchroom where the praise note intervention took place and observers confirmed there were zero instances of praise. Madsen et al. (1968) stated during baseline, there was no attempt to influence teacher behavior and teachers conducted class in their usual manner. Successful SCRDs for QI 6.5 that allowed for three possible demonstrations of experimental effect included multiple baseline and one A-B-A-B with maintenance (Hollingshead et al., 2016). Wheatley et al. (2009), for example, employed a multiple baseline design across three lunchroom behaviors (littering, inappropriate sitting, and running) with the whole school as the unit of analysis.

7.0. Outcome measures/DVs. All studies met QIs 7.1, 7.2, 7.3, and 7.5 for having reported socially valid outcomes, defined and described measurement of DVs, reported effects for all measures, and reported IOA ≥ 80%. Four studies (66.67%) met QI 7.4, frequency and timing of outcome measures, and QI 7.6 was not applicable to SCRDs. Only Haydon and Musti-Rao (2011) reported a specific social validity measure—teachers were interviewed about their perception of BSP delineated as three categories, ease of implementation, effectiveness, and likelihood of future use. Teachers reported BSP was not at all difficult to implement, liked it was only one thing to do and did not interrupt instruction, were pleased with the effect on student behavior, and might be very likely to use BSP in the future. Three studies defined on- or off-task behavior as the DV, two defined inappropriate or disruptive behavior as the DV, and Wheatley et al. (2009) measured littering, inappropriate sitting, and running in the lunchroom. Haydon and Musti-Rao (2011), as an example of QI 7.5, had two trained secondary observers record data simultaneous to the primary observer for 33% and 25% of sessions across phases, reporting mean IOA and ranges.

8.0. Data analysis. All studies met QI 8.2 with graphed outcome data for all phases and units of analysis. All studies also reported DV and IV data in tables or text allowing descriptive statistic comparison in addition to visual analysis. For example, the PE teacher in van der Mars (1989) increased BSP for three target students from baseline rates of 0.40, 0.36, and 0.44, respectively, to 1.06 (265% increase), 0.90 (250%), and 0.69 (156%) during intervention, and student off-task behavior dropped from 26.5- 39.4% of intervals to 9.3-15.4% of intervals.

Evidence Base Supporting BSP

We followed CEC (2014) standards for classifying the evidence base of BSP. First, we determined all included studies were methodologically sound by applying an 80% criterion (Lane et al., 2009) and eligible for classifying the evidence base of BSP (see Figure 2). The six articles met 87.13% of QIs or more, with weighted totals from 6.97 to 7.50 out of 8.00 QIs.

Next, we classified the six articles, all of which were SCRDs, as having *positive effects, neutral or mixed effects*, or *negative effects*. Two (van der Mars, 1989; Wheatley et al., 2009) established a functional relation between introduction of the IV and changes in DV outcomes (Horner & Odom, 2014), had three or more cases, and had at least 75% of cases with a meaningful therapeutic change, which we categorized as having *positive effects*. Two studies had less than three participants (Haydon & Musti-Rao, 2011; Hollingshead et al., 2016), and two studies used SCRDs without replications of effects needed to establish a functional relation (Houghton et al., 1990; Madsen et al., 1968). With two SCRDs with *positive effects*, we classified BSP in traditional K–12 settings as a *potentially EBP*.

Data Extraction and Analysis

Three indicators explored the effectiveness of teacher-delivered BSP studies (see Table 2). PND ranged from 41.18% to 100% (M = 83.26%, SD = 22.21) with three of six studies considered very effective, two effective, and one ineffective. When calculating Tau-U effect sizes, Houghton et al. (1990) and Haydon and Musti-Rao (2011) required

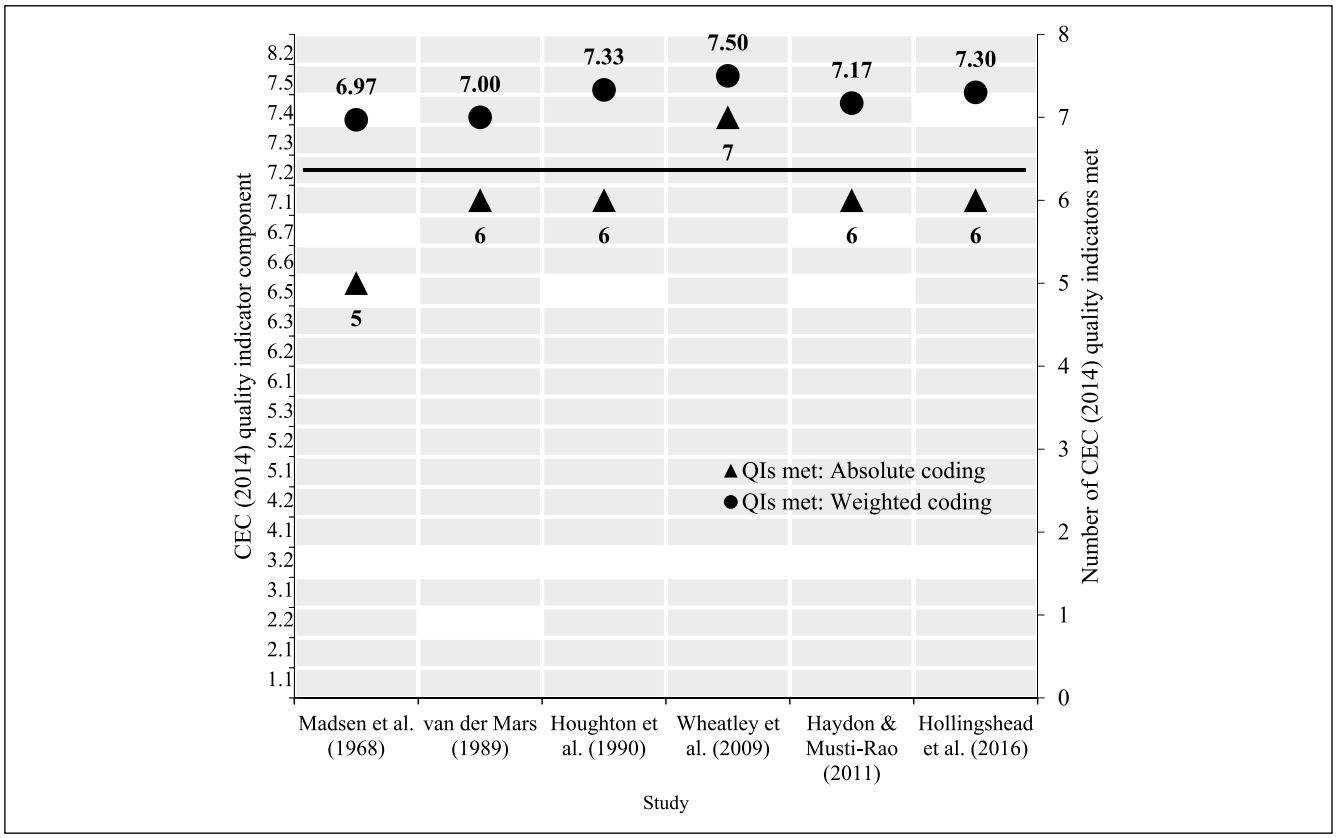

Figure 2. Teacher-delivered behavior-specific praise studies (abscissa) and CEC (2014) QI components met (primary ordinate; shaded cells = component met, white cells = component not met).

Note. Secondary ordinate (right *y* axis) displays number of QIs met by absolute coding (triangles; 8.0 QIs required) and weighted coding (circles), which required 6.4 QIs (80%) to be considered a methodologically sound study. The weighted coding criterion of 6.4 is indicated by the horizontal black line. CEC = Council for Exceptional Children; QI = quality indicator.

Table 2. Indicators of Teacher-Delivered Behavior-Specific Praise Study Effects.

				BC-SMD			
Study	Outcome measure	PND	Tau-U	Estimate	*SE*	95% CI range	EBP effects classification
Madsen, Becker, and Thomas (1968)	Inappropriate behavior	83.93	.94	1.77	0.59	[0.86, 3.01]	NA (functional relation not established)
van der Mars (1989)	Off-task behavior	94.44	.97	2.19	0.46	[1.34, 3.12]	Positive effects
Houghton, Wheldall, Jukes, and Sharpe (1990)	On-task behavior	80.00	.74	1.63	0.44	[0.80, 2.51]	NA (functional relation not established)
Wheatley et al. (2009)	Inappropriate littering, sitting, and running behaviors	100	1.00	1.42	1.14	[0.48, 3.51]	Positive effects
Haydon and Musti-Rao (2011)	Disruptions	100	.91	3.01	1.94	[1.09, 6.84]	NA (<3 cases; functional relation not established)
Hollingshead, Kroeger, Altus, and Trytten (2016)	On-task behavior	41.18	.73	—[a]	—	—	NA (<3 cases)

Note. PND = percentage of nonoverlapping data; BC-SMD = between-case standardized mean difference (Valentine, Tanner-Smith, Pustejovsky, & Lau, 2016); CI = confidence interval; EBP = evidence-based practice; NA = studies did not meet weighted criteria, had fewer than three cases reported, or did not demonstrate a functional relation between the introduction of the intervention and changes in outcome measures.

[a]Study did not meet the three-participant minimum criterion to be able to calculate BC-SMD.

baseline correction due to therapeutic baseline trend exceeding 20% of data pairs. Omnibus Tau-U effect sizes calculated for included articles ranged from .73 to 1.00 (M = 0.88, SD = 0.12). BC-SMD effect sizes for the five studies with three or more cases ranged from 1.42 to 3.01, all large effects. van der Mars (1989) and Wheatley et al. (2009), studies classified as having *positive effects* according to CEC (2014) standards, had large or very large effects on all three indicators. Although remaining studies demonstrated positive outcomes, they did not establish a functional relation between IV and DV or had less than three cases, so outcomes should be interpreted with caution.

Discussion

As more schools shift away from reactive discipline to a focus on preventing problem behavior through proactive positive behavioral supports and interventions, it becomes essential for teachers to be equipped with low-intensity classroom management strategies easily intensified to Tier 2 interventions (Lane, Menzies, et al., 2015). The classroom management practices literature review by Simonsen et al. (2008) identified specific contingent praise as effective and research supported, whereas Cavanaugh (2013) identified strategies to increase teachers' use of BSP. Combined with this systematic review, the literature shows BSP can be used by teachers for all students at each level of prevention in tiered models as this strategy is theoretically grounded in ABA (Cooper et al., 2007). Despite the perceived ease of using specific praise, natural rates of teacher BSP remain low, 1.75 per hour in Grades 1 to 6 compared with general praise at 29 per hour (Burnett & Mandel, 2010). Even when training on BSP is provided to teachers, improved rates of BSP do not always maintain over time (e.g., Hawkins & Heflin, 2011). Perhaps some teachers find it takes too much effort or extra time to make praise specific and varied instead of general, or lack knowledge of the ABA foundation behind using praise as reinforcement. Future studies could explore qualitatively why natural rates of BSP are low and why increased rates of BSP regress over time. Also, researchers might study quantitatively whether providing professional learning on the basics of ABA to preservice and/or inservice teachers increases and maintains higher rates of BSP, and how student outcomes are affected.

We focused our review on studies where the IV was teacher-implemented BSP and DVs included at least one student behavior outcome (e.g., on-task behavior, disruptive or inappropriate behavior), and found all studies demonstrated BSP as a simple, effective strategy to reduce undesirable behavior and promote appropriate behavior. Although only two studies were eligible for use in the classification of the evidence base for BSP, all studies were methodologically sound, demonstrating the challenge researchers continue to face designing and implementing rigorous experimental studies in applied settings such as traditional K–12 classrooms. For example, students might become ill or move half way through a well-designed intervention, or teacher attrition may occur. Life events are more likely to affect research when studies take place in traditional school settings compared with clinics or highly controlled settings.

It was encouraging to find studies published since 2009 reported participant race/ethnicity. Such information aids generalizability of BSP to multiple populations, as some studies took place in schools where the population was diverse (Haydon & Musti-Rao, 2011; Hollingshead et al., 2016). This illustrates the need to continue experimental research with BSP to grow the available evidence base and determine for whom and on what outcome measures EBPs work, so we may answer the critical question of external validity if we are to promote the use of BSP (Klingner & Edwards, 2006; Sloane, 2008; West et al., 2016).

In our coding of QIs, there were areas where we would have liked studies to report additional information even though not required to meet a particular QI. For example, QI 2.2 *participant risk status* was often not directly applicable to our review when the unit of analysis was the whole classroom. Nonetheless, it was helpful when studies reported how the class was selected for participation, or how direct observation or systematic processes occurred to confirm existence of a sufficient rate of target variables warranting intervention.

No studies reported setting a criterion for intervention agents to meet to demonstrate understanding of BSP before implementation in classrooms (QI 3.2). We believe this is an important QI to ensure treatment fidelity, as well-trained teachers would be more likely to follow intervention procedures and more likely to have greater competency delivering the intervention (Borrelli, 2011). Training to a criterion and checks for understanding, including role-play, provide opportunities for feedback, which have been shown to significantly increase implementers' skill acquisition and implementation performance (Roscoe & Fisher, 2008). A well-designed study yielding nonsignificant results may be due to an ineffective intervention or to lack of intervention agent training, stressing the need to check for understanding to a specific criterion when training intervention agents (Chambless & Hollon, 1998).

We anticipate some researchers and teachers might say this training QI was not relevant to our review. Some might state BSP is a simple concept not necessitating training or when training is provided, there is no need to check for understanding due to the intervention's simplicity. We required a check for understanding and criterion be met to ensure intervention tactics were well defined and understood by the treatment agent, as Haydon and Musti-Rao (2011) reported praising students does not always come naturally. Using BSP might especially feel awkward or

insincere at first when a teacher is not already used to delivering general praise, thus necessitating practice to make praise specific, authentic, and incorporate it into everyday teaching (Lane, Menzies, Bruhn, & Crnobori, 2011). Brophy (1981) reported praise is seldom used for good academic work, rarely used for good conduct, often used incorrectly, and often used not as deliberate reinforcement but when prompted by students seeking praise. Although recent studies have shown higher classroom praise rates (Burnett & Mandel, 2010; Jenkins et al., 2015), praise cannot be assumed to be reinforcing due to often improper and ineffective usage. To be reinforcing, praise must be contingent on student behavior, specific, sincere, varied, and credible, making training with checks for understanding essential (O'Leary & O'Leary, 1977).

We also would have liked authors to report more detail for QI 6.2 *describing baseline*. We only required authors to report data for teacher use of BSP (the IV) and student DVs as long as one general baseline characteristic was described (e.g., observations occurred during math, during group station work, during independent work while the teacher roamed). We preferred to know what typical routines were in place prior to intervention, as well as what school-wide structures were being designed or already in place as part of responsiveness to intervention (RtI; Fuchs & Fuchs, 2016), PBIS, or Ci3T, if applicable. Adding rich description to baseline is essential for the reader to be able to evaluate IV effects, generalize results, and inform replications (Lane, Wolery, Reichow, & Rogers, 2007).

With five out of six included studies demonstrating large or very large effects on three indicators (PND, Tau-U, BC-SMD), BSP appears to be an effective strategy for increasing desired academic and social behavior while decreasing problem behaviors. Generalizability of these effects will require additional studies, however, to determine for whom BSP works, when, and in what situations (Wolery & Dunlap, 2001).

Although all studies demonstrated positive outcomes for all participants, including increased on-task behavior, increased punctuality, and decreased inappropriate and problem behaviors, only two were able to be categorized as having *positive effects* by CEC (2014) standards. Other studies met the 80% weighted criterion but did not meet *positive effects* criteria due to having less than three participants or not demonstrating a functional relation between the IV and DVs. This does not diminish the positive benefits of BSP demonstrated by these studies. Rather, it provides direction for future researchers. Specifically, we need additional, rigorously designed and executed studies to better understand the utility of BSP with various populations, student outcome measures, and contexts. In other words, how well does BSP work to improve Y (DVs) for P (participants) in S (setting; Kettler & Lane, 2016)?

Limitations and Future Directions

Interpretation of our findings should be made with caution given the following limitations. First, we focused on teacher-delivered BSP as the IV, warranted by the variety of IVs encountered (i.e., teacher-delivered BSP, coaching of teachers to increase use of BSP, peer-implemented BSP, researcher- or experimenter-delivered BSP), which excluded additional studies. We recommend future inquiry investigate coaching of teacher BSP and peer-delivered BSP to establish whether either can be considered an EBP.

A second limitation is the interpretability of QIs. Although we used systematic training procedures to align our understanding for article selection and coding, ambiguity remains, made evident by the range of IRA percentages across search steps (e.g., 75.00% for ancestral abstracts) and QI component coding (50%–100%). Low IRA during ancestral abstract reading was due to readers differing in their over- or underinclusion of articles when DVs were not clear or praise statements were not clearly behavior specific in the limited space of the abstract, whereas low IRA between QI components was due to having only six included articles. Specifically, each reader had a different interpretation on what constituted a check for understanding in QI 3.2 *intervention agent training*, for QI 2.2 *disability or risk status* when the whole class was the unit of analysis, and for QI 2.2, if teacher nomination of students required additional observation to confirm high rates of behavior. This was resolved after each discrepancy first arose, but with only six articles, there were not enough studies to apply consensus understanding and raise IRA. The variability we found in interpretation of CEC (2014) QIs sends a stronger call to the research community to design and report studies with greater rigor, specificity, and attention to scientific methods. We hope our colleagues will heed this call and conduct additional studies on BSP that establish a functional relation between IV and DVs, including randomized or quasi-experimental group studies if possible, to further add to the evidence base for BSP.

It should also be noted for our inclusion of Madsen et al. (1968), authors collected data in categories for verbal and nonverbal praise (e.g., smiles, nods, pats on the shoulder), but in reporting, those data were collapsed into one teacher approval category. We included the study because Madsen and colleagues stated 85% of the category comprised verbal praise, though a functional relation was not established and the article was not used in the classification of the evidence base for BSP. We advise future studies to report data using categories they were collected in, with additional merged categories as needed to answer research questions.

As researchers work to add to the evidence base for BSP, we encourage the field to design studies allowing for a functional relation. For example, when using multiple baseline designs across teachers (e.g., Haydon & Musti-Rao,

2011), research teams could recruit four teacher participants in case one takes a leave of absence or otherwise withdraws from the study, allowing three remaining teachers to demonstrate a functional relation between the IV and DVs (Gast & Ledford, 2014). Similarly, if it is desired to switch the order of interventions in a multiple baseline design (e.g., Houghton et al., 1990), we encourage researchers to plan four rows or tiers to each set of ordered interventions (e.g., four teachers following an A-B-C multiple baseline design and four ordered as A-C-B), so both sets of intervention order have the opportunity to demonstrate an experimental effect with at least two replications (Gast & Ledford, 2014) even if one teacher from each set withdraws from the study.

We also recommend journal editors require authors report detailed context and setting descriptions and background variables for intervention agents, or a justification as to why these data are not publishable. In addition, we recommend researchers report participant demographics in SCRDs even though they may not always be considered necessary (West et al., 2016; Wolery & Ezell, 1993). Finally, we recommend future studies include rich descriptions of BSP statements along with implementation fidelity data collected on the consistency of statements used, allowing for QI coding and comparison across studies.

Educational Implications

As a *potentially EBP*, it will be necessary to establish additional evidence for the effectiveness of BSP in traditional K–12 settings with various populations before any formal recommendations or generalization can be made. Wheatley et al. (2009) demonstrated the effectiveness of a simple lunchroom praise note system paired with verbal BSP, showing undesirable behaviors (littering, sitting inappropriately, and running) can be quickly and dramatically reduced by explicitly stating expectations, providing practice opportunities, and reinforcing students who meet expectations with BSP (verbal and written). Interested schools might consider replicating this intervention design for similar lunchroom concerns.

The studies in this review involved teacher use of BSP delivered to all students in the classroom contingent on meeting prescribed expectations. Collectively, these studies demonstrated versatility of BSP, as interventions occurred in PE, kindergarten, second grade, middle school math, seventh-grade social studies, and in England elementary English, geography, math, and chemistry classes, all with positive student outcomes. We encourage all teachers to use BSP as part of Tier 1 everyday practice to establish and sustain a positive, safe learning environment following PBIS and ABA tenets. When classroom behavior concerns arise, teachers might consider following procedures outlined in the reviewed BSP studies to target students with behavior concerns. In addition, we refer interested readers to *Supporting Behavior for School Success: A Step-by-Step Guide to Key Strategies* (Lane, Menzies, et al., 2015), which includes seven steps for implementing BSP with fidelity. Teachers can self-monitor use of BSP (e.g., MotivAider in Haydon & Musti-Rao, 2011; hand tally counter in Houghton et al., 1990) without disruption to teaching procedures and collect permanent product outcomes (e.g., assignment completion rate, assignment accuracy) or student behavior data (e.g., on-task behavior, disruptive behavior). As needed, paraeducators, coteachers, or other staff can help collect observation data (e.g., Lane, Royer, et al., 2015), or teachers can monitor student behavior and collect data on their own. For example, moving buttons or paperclips from one pocket to the other for frequency counts (Freeman & Tieghi-Benet, n.d.; Oklahoma State Department of Education—Special Education Services, 2014), taking tally marks on a clipboard or address label sticker placed on clothing (e.g., upper thigh), or wearing a vibrating device (e.g., MotivAider) set at a predetermined interval to signal when to observe the target student and mark if he or she is on task. Future researchers might review the literature to determine the extent to which various methods have been successful in coaching teachers to increase and sustain higher rates of BSP in their classroom. In short, with a little practice, teachers can use BSP to reduce problem behavior and increase on-task behavior of targeted students or their whole class.

Summary

BSP is a *potentially EBP* for use at all tiers to support positive classroom climates full of warmth. Teachers can use BSP daily as one form of acknowledging students who meet defined academic, behavior, and social expectations. BSP can also be used as a low-intensity Tier 1 teacher-delivered strategy to prevent and reduce disruptive, inappropriate, or problem behavior, and as a Tier 2 strategy to target an individual or small group of students with academic, behavior, or social concerns. The six studies reviewed show BSP can be implemented with minimal effort and virtually no disruption to teaching routines, prompting quick changes in student outcome measures. More studies are needed with high methodological rigor to determine whether BSP in K–12 traditional school settings can be considered an *EBP*.

Declaration of Conflicting Interests

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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDS

David J. Royer https://orcid.org/0000-0003-2882-1049

Kathleen Lynne Lane's https://orcid.org/0000-0001-6364-838X

References

References with an asterisk (*) indicate studies included in the systematic review.

Allday, R. A., Hinkson-Lee, K., Hudson, T., Neilsen-Gatti, S., Kleinke, A., & Russel, C. S. (2012). Training general educators to increase behavior-specific praise: Effects on students with EBD. *Behavioral Disorders*, *37*, 87–98.

Beaman, R., & Wheldall, K. (2000). Teachers' use of approval and disapproval in the classroom. *Educational Psychology*, *20*, 431–446. doi:10.1080/713663753

Borrelli, B. (2011). The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. *Journal of Public Health Dentistry*, *71*(S1), S52–S63. doi:10.1111/j.1752-7325.2011.00233.x

Brophy, J. (1981). Teacher praise: A functional analysis. *Review of Educational Research*, *51*, 5–32. doi:10.2307/1170249

Burnett, P. C., & Mandel, V. (2010). Praise and feedback in the primary classroom: Teachers' and students' perspectives. *Australian Journal of Educational & Developmental Psychology*, *10*, 145–154.

Caldarella, P., Christensen, L., Young, K. R., & Densley, C. (2011). Decreasing tardiness in elementary school students using teacher-written praise notes. *Intervention in School and Clinic*, *47*, 104–112. doi:10.1177/1053451211414186

Cavanaugh, B. (2013). Performance feedback and teachers' use of praise and opportunities to respond: A review of the literature. *Education and Treatment of Children*, *36*, 111–136.

Chalk, K., & Bizo, L. A. (2004). Specific praise improves on-task behaviour and numeracy enjoyment: A study of year four pupils engaged in the numeracy hour. *Educational Psychology in Practice*, *20*, 335–351. doi:10.1080/0266736042000314277

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *66*, 7–18. doi:10.1037//0022-006x.66.1.7

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson.

Council for Exceptional Children. (2014). *Council for Exceptional Children standards for evidence-based practices in special education*. Arlington, VA: Author.

Drevon, D., Fursa, S. R., & Malcolm, A. L. (2016). Intercoder reliability and validity of WebPlotDigitizer in extracting graphed data. *Behavior Modification*, *41*, 1–17. doi:10.1177/0145445516673998

Everett, G. E., Olmi, D. J., Edwards, R. P., & Tingstrom, D. H. (2005). The contributions of eye contact and contingent praise to effective instruction delivery in compliance training. *Education and Treatment of Children*, *28*, 48–62.

Floress, M. T., & Jenkins, L. N. (2015). A preliminary investigation of kindergarten teachers' use of praise in general education classrooms. *Preventing School Failure: Alternative Education for Children and Youth*, *59*, 253–262. doi:10.1080/1045988X.2014.942834

Freeman, R., & Tieghi-Benet, M. (n.d.). *Data-based decision making*. Retrieved from http://www.specialconnections.ku.edu/~kucrl/cgi-bin/drupal/?q=assessment/data_based_decision_making

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*, 2–18. doi:10.1037/a0024338

Fuchs, D., & Fuchs, L. S. (2016). Responsiveness-to-intervention: A "systems" approach to instructional adaptation. *Theory Into Practice*, *55*, 225–233. doi:10.1080/00405841.2016.1184536

Gable, R. A., Hendrickson, J. M., Young, C. C., Shores, R. E., & Stowitschek, J. J. (1982). A comparison of teacher approval and disapproval statements across categories of exceptionality. *Journal of Special Education Technology*, *6*, 15–22. doi:10.1177/016264348300600103

Gast, D. L., & Ledford, J. R. (Eds.). (2014). *Single case research methodology: Applications in special education and behavioral sciences* (2nd ed.). New York, NY: Routledge.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, *71*, 149–164. doi:10.1177/001440290507100202

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112. doi:10.3102/003465430298487

Hawkins, S. M., & Heflin, L. J. (2011). Increasing secondary teachers' behavior-specific praise using a video self-modeling and visual performance feedback intervention. *Journal of Positive Behavior Interventions*, *13*, 97–108. doi:10.1177/1098300709358110

*Haydon, T., & Musti-Rao, S. (2011). Effective use of behavior-specific praise: A middle school case study. *Beyond Behavior*, *20*(2), 31–39.

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods*, *3*, 224–239. doi:10.1002/jrsm.1052

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods*, *4*, 324–341. doi:10.1002/jrsm.1086

*Hollingshead, A., Kroeger, S. D., Altus, J., & Trytten, J. B. (2016). A case study of positive behavior supports-based interventions in a seventh-grade urban classroom. *Preventing School Failure: Alternative Education for Children and Youth*, *60*, 278–285. doi:10.1080/1045988X.2015.1124832

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*, 165–179. doi:10.1177/001440290507100203

Horner, R. H., & Odom, S. L. (2014). Constructing single-case research designs: Logic and options. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case intervention research: Methodological and statistical advances* (pp. 27–51). Washington, DC: American Psychological Association.

Horner, R. H., & Sugai, G. (2015). School-wide PBIS: An example of applied behavior analysis implemented at a scale of social importance. *Behavior Analysis in Practice*, *8*, 80–85. doi:10.1007/s40617-015-0045-4

*Houghton, S., Wheldall, K., Jukes, R., & Sharpe, A. (1990). The effects of limited private reprimands and increased private praise on classroom behaviour in four British secondary school classes. *British Journal of Educational Psychology*, *60*, 255–265. doi:10.1111/j.2044-8279.1990.tb00943.x

Jenkins, L. N., Floress, M. T., & Reinke, W. (2015). Rates and types of teacher praise: A review and future directions. *Psychology in the Schools*, *52*, 463–476. doi:10.1002/pits.21835

Kennedy, C., & Jolivette, K. (2008). The effects of positive verbal reinforcement on the time spent outside the classroom for students with emotional and behavioral disorders in a residential setting. *Behavioral Disorders*, *33*, 211–221.

Kettler, R. J., & Lane, K. L. (2016). *Methodological foundations of school psychology research and practice*. New York, NY: Routledge.

Klingner, J. K., & Edwards, P. A. (2006). Cultural considerations with response to intervention models. *Reading Research Quarterly*, *41*, 108–117. doi:1598/RRQ.41.1.6

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Lane, K. L., Common, E. A., Royer, D. J., & Muller, K. (2014). *Council for Exceptional Children 2014 group comparison and single-case research design standards quality indicator matrix*. Unpublished tool. Available from http://www.ci3t.org/practice

Lane, K. L., Kalberg, J. R., & Shepcaro, J. C. (2009). An examination of the evidence base for function-based interventions for students with emotional and/or behavioral disorders attending middle and high schools. *Exceptional Children*, *75*, 321–340. doi:10.1177/001440290907500304

Lane, K. L., Menzies, H. M., Bruhn, A. L., & Crnobori, M. E. (2011). *Managing challenging behaviors in schools: Research-based strategies that work*. New York, NY: Guilford Press.

Lane, K. L., Menzies, H. M., Ennis, R. P., & Oakes, W. P. (2015). *Supporting behavior for school success: A step-by-step guide to key strategies*. New York, NY: Guilford Press.

Lane, K. L., Oakes, W. P., & Menzies, H. M. (2014). Comprehensive, integrated, three-tiered models of prevention: Why does my school—and district—need an integrated approach to meet students' academic, behavioral, and social needs? *Preventing School Failure: Alternative Education for Children and Youth*, *58*, 121–128. doi:10.1080/10459 88X.2014.893977

Lane, K. L., Royer, D. J., Messenger, M. L., Common, E. A., Ennis, R. P., & Cantwell, E. D. (2015). Empowering teachers with low-intensity strategies to support academic engagement: Implementation and effects of instructional choice for elementary students in inclusive settings. *Education and Treatment of Children*, *38*, 473–504. doi:10.1353/etc.2015.0013

Lane, K. L., & Walker, H. M. (2015). The connection between assessment and intervention: How does screening lead to better interventions? In B. Bateman, J. W. Lloyd & M. Tankersley (Eds.), *Enduring issues in special education: Personal perspectives* (pp. 283–301). New York, NY: Routledge.

Lane, K. L., Wolery, M., Reichow, B., & Rogers, L. (2007). Describing baseline conditions: Suggestions for study reports. *Journal of Behavioral Education*, *16*, 224–234. doi:10.1007/s10864-006-9036-4

*Madsen, C. H., Becker, W. C., & Thomas, D. R. (1968). Rules, praise, and ignoring: Elements of elementary classroom control. *Journal of Applied Behavior Analysis*, *1*, 139–150. doi:10.1901/jaba.1968.1-139

Nickerson, A. B., & Brosof, A. M. (2003). Identifying skills and behaviors for successful inclusion of students with emotional or behavioral disorders. *Behavioral Disorders*, *28*, 401–409.

Oakes, W. P., Lane, K. L., & Germer, K. A. (2014). Developing the capacity to implement tier 2 and tier 3 supports: How do we support our faculty and staff in preparing for sustainability? *Preventing School Failure: Alternative Education for Children and Youth*, *58*, 183–190. doi:10.1080/10459 88X.2014.895575

Oklahoma State Department of Education—Special Education Services. (2014). *Behavioral interventions: Teacher self-assessment instructions*. Available from http://sde.ok.gov/

O'Leary, K. D., & O'Leary, S. G. (Eds.). (1977). *Classroom management: The successful use of behavior modification* (2nd ed.). New York, NY: Pergamon Press.

Oliver, R. M., & Reschly, D. J. (2010). Special education teacher preparation in classroom management: Implications for students with emotional and behavioral disorders. *Behavioral Disorders*, *35*, 188–199.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy*, *42*, 284–299. doi:10.1016/j.beth.2010.08.006

Polick, A. S., Carr, J. E., Davis, J. L., & Hanney, N. M. (2012). A comparison of general and descriptive praise in teaching intraverbal behavior to children with autism. *Journal of Applied Behavior Analysis*, *45*, 593–599. doi:10.1901/jaba.2012.45-593

Pustejovsky, J. E. (2016a). *Tau-U*. Retrieved from http://jepusto.github.io/Tau-U

Pustejovsky, J. E. (2016b). *What is Tau-U?* Retrieved from http://jepusto.github.io/What-is-Tau-U

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

Resnik, D. B., & Elmore, S. A. (2016). Ensuring the quality, fairness, and integrity of journal peer review: A possible role of editors. *Science and Engineering Ethics*, *22*, 169–188. doi:10.1007/s11948-015-9625-5

Rohatgi, A. (2017). *WebPlotDigitizer* (version 3.11) [web-based application]. Retrieved from http://arohatgi.info/WebPlotDigitizer/

Roscoe, E. M., & Fisher, W. W. (2008). Evaluation of an efficient method for training staff to implement stimulus preference assessments. *Journal of Applied Behavior Analysis*, *41*, 249–254. doi:10.1901/jaba.2008.41-249

Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single-subject research: Issues and applications. *Behavior Modification*, *22*, 221–242. doi:10.1177/01454455980223001

Scruggs, T. E., & Mastropieri, M. A. (2013). PND at 25: Past, present, and future trends in summarizing single-subject research. *Remedial and Special Education*, *34*, 9–19. doi:10.1177/0741932512440730

Shores, R. E., Jack, S. L., Gunter, P. L., Ellis, D. N., DeBriere, T. J., & Wehby, J. H. (1993). Classroom interactions of children with behavior disorders. *Journal of Emotional and Behavioral Disorders*, *1*, 27–39. doi:10.1177/106342669300100106

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

Slaughter, S. E., Hill, J. N., & Snelgrove-Clarke, E. (2015). What is the extent and quality of documentation and reporting of fidelity to implementation strategies: A scoping review. *Implementation Science*, *10*, 129. doi:10.1186/s13012-015-0320-3

Sloane, F. (2008). Comments on Slavin: Through the looking glass: Experiments, quasi-experiments, and the medical model. *Educational Researcher*, *37*, 41–46. doi:10.3102/0013189X08314835

Smith, S. C., Lewis, T. J., & Stormont, M. (2011). The effectiveness of two universal behavioral supports for children with externalizing behavior in head start classrooms. *Journal of Positive Behavior Interventions*, *13*, 133–143. doi:10.1177/1098300710379053

Stormont, M., & Reinke, W. (2009). The importance of precorrective statements and behavior-specific praise and strategies to increase their use. *Beyond Behavior*, *18*(3), 26–32.

Sutherland, K. S., Wehby, J. H., & Copeland, S. R. (2000). Effect of varying rates of behavior-specific praise on the on-task behavior of students with EBD. *Journal of Emotional and Behavioral Disorders*, *8*, 2–8. doi:10.1177/106342660000800101

Thompson, M. T., Marchant, M., Anderson, D., Prater, M. A., & Gibb, G. (2012). Effects of tiered training on general educators' use of specific praise. *Education and Treatment of Children*, *35*, 521–546. doi:10.1353/etc.2012.0032

Valentine, J. C., Tanner-Smith, E. E., Pustejovsky, J. E., & Lau, T. S. (2016). *Between-case standardized mean difference effect sizes for single-case designs: A primer and tutorial using the scdhlm web application*. Oslo, Norway: The Campbell Collaboration. doi:10.4073/cmdp.2016.1

*van der Mars, H. (1989). Effects of specific verbal praise on off-task behavior of second-grade students in physical education. *Journal of Teaching in Physical Education*, *8*, 162–169. doi:10.1123/jtpe.8.2.162

Vannest, K. J., & Ninci, J. (2015). Evaluating intervention effects in single-case research designs. *Journal of Counseling & Development*, *93*, 403–411. doi:10.1002/jcad.12038

Vannest, K. J., Parker, R. I., Gonen, O., & Adiguzel, T. (2016). *Single case research: Web based calculators for SCR analysis* (version 2.0) [web-based application]. Available from http://www.singlecaseresearch.org/

Weaver, A. D., & Watson, T. S. (2004). An idiographic investigation of the effects of ability-and effort-based praise on math performance and persistence. *The Behavior Analyst Today*, *5*, 381–390. doi:10.1037/h0100045

Wehby, J. H., Symons, F. J., & Shores, R. E. (1995). A descriptive analysis of aggressive behavior in classrooms for children with emotional and behavioral disorders. *Behavioral Disorders*, *20*, 87–105.

West, E. A., Travers, J. C., Kemper, T. D., Liberty, L. M., Cote, D. L., McCollow, M. M., & Brusnahan, L. L. S. (2016). Racial and ethnic diversity of participants in research supporting evidence-based practices for learners with autism spectrum disorder. *The Journal of Special Education*, *50*, 151–163. doi:10.1177/0022466916632495

What Works Clearinghouse. (2013). *What Works Clearinghouse: Procedures and standards handbook* (version 3.0). Retrieved from http://ies.ed.gov/ncee/wwc/

*Wheatley, R. K., West, R. P., Charlton, C. T., Sanders, R. B., Smith, T. G., & Taylor, M. J. (2009). Improving behavior through differential reinforcement: A praise note system for elementary school students. *Education and Treatment of Children*, *32*, 551–571.

White, M. A. (1975). Natural rates of teacher approval and disapproval in the classroom. *Journal of Applied Behavior Analysis*, *8*, 367–372. doi:10.1901/jaba.1975.8-367

Wolery, M., & Dunlap, G. (2001). Reporting on studies using single-subject experimental methods. *Journal of Early Intervention*, *24*, 85–89. doi:10.1177/105381510102400201

Wolery, M., & Ezell, H. K. (1993). Subject descriptions and single-subject research. *Journal of Learning Disabilities*, *26*, 642–647. doi:10.1177/002221949302601001