



OPEN ACCESS

EDITED BY

Gregor Ross Maxwell,
UiT The Arctic University of Norway, Norway

REVIEWED BY

Rosario Ruiz-Olivares,
University of Córdoba, Spain
Vanessa Scherman,
International Baccalaureate (IBO),
Netherlands

*CORRESPONDENCE

David James Royer
✉ david.royer@louisville.edu

†PRESENT ADDRESS

Robin Parks Ennis,
Department of Special Education and Multiple
Abilities, University of Alabama,
Tuscaloosa, AL, United States

RECEIVED 05 June 2024

ACCEPTED 01 August 2024

PUBLISHED 13 September 2024

CITATION

Royer DJ and Ennis RP (2024)
Student-delivered behavior-specific praise: a
systematic literature review and
meta-analysis.
Front. Educ. 9:1444394.
doi: 10.3389/feeduc.2024.1444394

COPYRIGHT

© 2024 Royer and Ennis. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Student-delivered behavior-specific praise: a systematic literature review and meta-analysis

David James Royer^{1*} and Robin Parks Ennis^{2†}

¹Department of Special Education, Early Childhood, and Prevention Science, University of Louisville, Louisville, KY, United States, ²Department of Curriculum and Instruction, University of Alabama at Birmingham, Birmingham, AL, United States

Behavior-specific praise is an easy-to-implement, teacher-delivered strategy that supports academic engagement while preventing and reducing disruptive behavior. By letting students know what they did, specifically, to meet academic, behavioral, and/or social expectations, students who find teacher attention reinforcing are more likely to engage in the same behavior more often in the future. While teacher-delivered behavior-specific praise was classified as a *potentially evidence-based practice* using Council for Exceptional Children standards, less is known about the effects of students who deliver behavior-specific praise to their peers. This systematic literature review and meta-analysis explored the literature base and found 36 articles meeting inclusion criteria. Fifteen articles included positive peer reporting as the independent variable, 20 included tootling as the intervention, two compared those interventions, and three used an “other” form of peer praise (i.e., peer praise notes, peer monitor tokens). Nine tootling articles met all eight quality indicators by absolute coding, and 32 out of all 36 studies met an 80% weighted quality indicator coding criterion for being methodologically sound. From these, we classified positive peer reporting in the *mixed evidence* category and tootling in the *evidence-based practice* category. We discuss benefits of various components in each type of peer praise intervention, limitations of the literature review, and make recommendations for future researchers.

KEYWORDS

behavior-specific praise, literature review, meta-analysis, tootling, positive peer reporting

Introduction

Students with emotional and behavioral disorders (EBD) are those who have difficulty meeting school expectations, from following the rules, to performing academically at grade level, to sustaining appropriate peer and adult relationships (Mundschen and Simpson, 2014). Point prevalence estimates indicate 12% of students have at least a moderate EBD and 20% have at least a mild EBD (Forness et al., 2012), yet only 0.5% of students received special education services under the emotional disturbance (ED) category of *Individuals with Disabilities Education Improvement Act (2004)* each year from 2011 to 2020 (latest data), down from 0.7% in years 2005–2007 and 0.6% in years 2008–2010 (U.S. Department of Education, 2022). This means most students with EBD, those classified as having externalizing (e.g., aggression, defiance, arguing, disruptive behavior, rule violations, substance use; Romer et al., 2020) and/or internalizing (e.g., withdrawal, negative affect, anxiety, depression; Romer et al., 2020)

behavior patterns, attend general education classes and do not receive special education support.

As a result of the challenges associated with EBD, students with or at risk for EBD often experience social isolation, peer rejection, and fewer positive interactions with students and adults (Zweers et al., 2021). Certainly this makes sense, as young students especially may not have the social skills to develop a good relationship with someone who is volatile, or they may not want to be friends with a peer they perceive to frequently get in trouble at school. Similarly, for adults, without the skills and strategies needed to support students with EBD, it can be difficult for teachers to maintain a positive or supportive relationship with them (O'Connor et al., 2011), and it may seem easier for some teachers to simply send a student with EBD out of the classroom when they are repeatedly disruptive, for example. Over the last 20 years or so, however, more schools are working to adopt tiered models that prevent most challenging behavior, such as positive behavioral interventions and supports (PBIS; Sugai and Horner, 2019) and the comprehensive, integrated, three-tiered (Ci3T; Lane et al., 2019b) model of prevention. Within these models of increasingly intensive student supports, educators are empowered with tools and low-intensity strategies that increase their classroom self-efficacy and give them confidence in their ability to keep students with challenging behavior in the classroom learning.

One way for teachers to increase positive interactions with all students at Tier 1, including those with or at risk for EBD, is to focus on the ratio of positive statements to corrections and reprimands (Caldarella et al., 2023). Many teachers receive training at some point early in their career about having a 4:1 or 5:1 ratio of positive to negative statements, such as learning for every academic correction (e.g., “you forgot to take the reciprocal of the fraction”) to also give a few positive acknowledgments (e.g., “I see you were following the mnemonic we learned yesterday and I like that you remembered to isolate the variable”) immediately, and more later even if unrelated to the initial correction (e.g., “Your printing is very neat,” “Thank you for raising your hand and waiting quietly,” “Well done”) in order to get to a higher ratio of positives to negatives. These positive statements can have the most impact on future student academic performance and behavior when they are specific in identifying exactly what the student did well (Brophy, 1981).

Behavior-specific praise (BSP) is a form of positive reinforcement that specifically acknowledges desired behaviors and strengthens the likelihood socially acceptable behaviors will occur more often in the future, especially when students like the attention (Cooper et al., 2020). BSP statements can be written or oral, indicating precisely the behavior observed (including academic behavior) that met expectations (Menzies et al., 2023). For example, a teacher might say to on-task students during math work, “I like the way you are using your small white board to show your math work” when there was a past incident of inappropriate white board use or to promote continued appropriate use. When BSP is sincere, varied, targets effort instead of ability, and the student finds attention reinforcing, what was specifically praised is more likely to occur more often (Lane et al., 2015). This contrasts with general praise, where a specific action is not identified, such as saying, “Good job” or giving a thumbs up with a smile. General praise is a good way to increase positive interactions with students too, but BSP has the added benefit of specifying exactly what expectations were met, not only for the child receiving the praise but as a reminder to all students nearby (Sutherland et al., 2000). There are many studies showing the impact BSP has on increasing academic

engagement and decreasing disruptive behavior, and the strategy has been classified as a *potentially evidence-based practice* applying Council for Exceptional Children (2014) standards (Royer et al., 2019).

Unfortunately, naturally occurring rates of positive feedback are “alarmingly low” (Scott et al., 2017, p. 61), even at the elementary level where the rate per minute is 0.137 on average, with a positive to negative ratio of 3:1. That means only every 7.5 min does an elementary student typically receive positive feedback from their teacher, and the rates and ratios are even lower at middle school (0.061 or every 16.4 min; 1.74:1) and high school (0.033 or every 30.3 min; 0.65:1; Scott et al., 2017). Obviously, there is a need for students to receive higher rates of positive interactions from their teachers, and potentially, their peers can help as well.

Student-delivered behavior-specific praise

Students can deliver praise to their peers and help increase the number of positive interactions, especially for students with EBD and/or who are socially isolated (usually due to internalizing behavior patterns) or socially rejected (usually due to externalizing behavior patterns). Collins et al. (2020) conducted a meta-analytic review of peer-reporting interventions utilizing single-case research designs and identified 21 studies meeting inclusion criteria. Their findings suggested peer reporting interventions had a positive impact on student behavior outcomes, noting variability among included studies' approaches to peer praise. Additionally, authors compared studies using log response ratios, tau (measure of overlap), and moderating effects of targeted contextual variables. While Collins et al. (2020) applied elements of What Works Clearinghouse standards to their inclusion criteria, they did not code studies for quality indicators and did not evaluate each approach to peer praise in isolation. When Ennis et al. (2020) mapped the 50-year knowledge base on BSP, they found six journal articles on peer praise, such as peer praise notes used to increase social interactions among three junior high school students at risk for EBD (Peterson Nelson et al., 2008) and peer praise notes to reduce problem behaviors at recess for an elementary school with 462 students (Teerlink et al., 2017). Even more prolific than praise notes were the approaches to peer praise called positive peer reporting and tootling.

Positive peer reporting

Positive peer reporting (PPR) is a brief period of time for peers to publicly praise typically one “star” target student with BSP, encouraging prosocial behavior and earning tokens for each appropriate BSP toward a class reward (group contingency). PPR interventions are generally designed to increase the frequency and improve the quality of the target student's prosocial interactions with peers (Morrison and Jones, 2007) and have added benefits for the whole class's behavior given the group contingency. When the star is not known (one variation of PPR), students, in theory, are on their best behavior in case they will be the ones publicly praised later. Studies vary in terms of how long a student was the star (e.g., changed each day, each week), how many stars (e.g., one, three), when peers observe the star for prosocial behaviors (e.g., all day, during one subject), if the star is known or unknown, and when and for how long peers publicly praise the star (e.g., end of subject for 3 min, end of day

for 10 min). All but one PPR study was published before 2014, when the peer praise literature turned all but exclusively to investigating tootling interventions.

Tootling

Tootling is a classwide application of PPR where students observe all peers instead of one or a few stars and privately report specific prosocial behaviors on index cards to the teacher. Each appropriate tootle with required components (e.g., name of both students giving and receiving praise, praise statement is specific) earns points toward a class reward (group contingency). The name tootling comes from merging ‘toot your own horn’ and tattling and is intended to be the opposite of tattling (Skinner et al., 1998). Tootling interventions vary in terms of how long of a time period peer observations occur (e.g., all day, during one subject), how many tootles can be written and turned in (e.g., two maximum per session, unlimited), and when and for how long the teacher reads tootles to the class (e.g., end of subject for 5 min, end of day for 3 min). Some studies included a public posting of tootles for everyone to read, either using technology like Class Dojo for live display when entered by students electronically (McHugh Dillon et al., 2019) or using paper posted to a bulletin board (Harry et al., 2023).

In a seemingly transitional time of researchers shifting focus from PPR to tootling, two studies compared PPR to tootling. Barahona (2010) found neither intervention reduced disruptive behavior more than a minimal amount across three elementary grade 3 general education classrooms, while in contrast, Sherman (2012) found both PPR and tootling increased appropriate behavior and reduced inappropriate behavior for four students in general education classrooms grades 3–6. More analyses are therefore needed to determine how PPR compares to tootling and how effective peer-delivered praise is, generally.

Purpose

Given the emphasis in PBIS and Ci3T tiered models of prevention on teachers using the low-intensity strategy of behavior-specific praise (BSP) to support positive, productive, safe learning environments, and given Ennis et al. (2020) found six peer praise studies but did not include theses and dissertations, the purpose of this systematic literature review and meta-analysis was to explore student-delivered praise further. Specifically, our research questions were: (a) To what extent did peer praise interventions address Council for Exceptional Children (2014) quality indicators of methodologically sound studies? (b) What is the evidence-based practice status of peer praise according to Council for Exceptional Children (2014) guidelines, applying an 80% minimum criterion for methodologically sound studies (Lane et al., 2009)? (c) What was the magnitude of effects for peer praise interventions?

Method

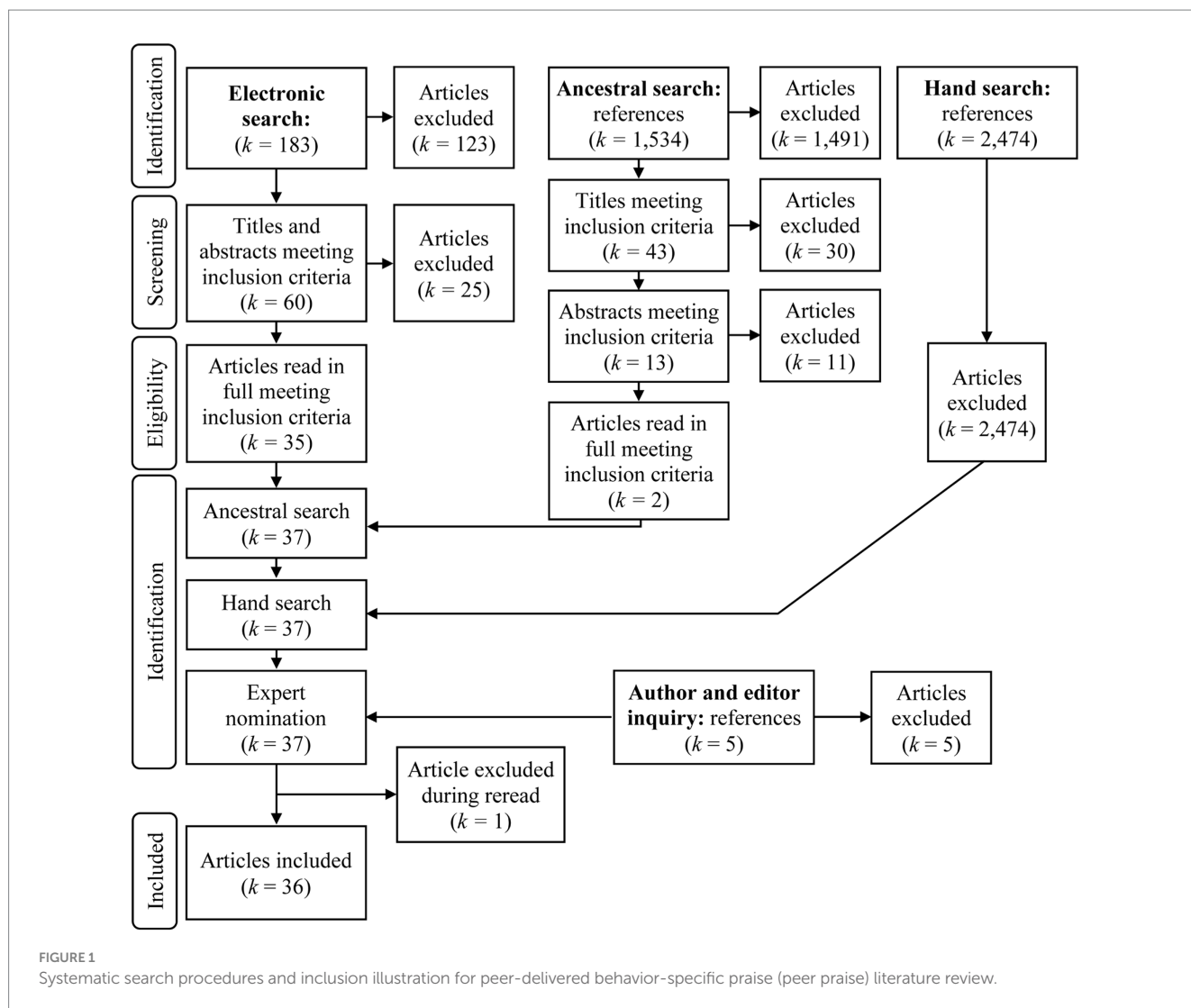
Search and article selection

We conducted an exhaustive search of student-delivered BSP research, involving four search steps: (1) electronic, (2) ancestral, (3)

hand, and (4) expert nomination (Lane et al., 2022). First, we searched Educational Resources Information Center (ERIC), ProQuest Dissertations and Theses Global, American Psychological Association (APA) PsycINFO, APA PsychARTICLES, and Research Library through December 2023 using Boolean search terms (behavio* AND specific AND praise AND peer) OR (tootling), “peer praise note*,” and “positive peer reporting.” This search returned 183 unique manuscripts (articles, theses/dissertations) after duplicates were removed (see Figure 1). Both authors independently screened titles and abstracts for inclusion, and interrater reliability (IRR) was 97.81% and Cohen’s $\kappa=0.95$ [95% CI=0.9, 1.0], which takes chance agreement into consideration, indicated near-perfect agreement (Cohen, 1960; Landis and Koch, 1977), resulting in 60 manuscripts to read in full. Both authors independently read in full and found 35 manuscripts for inclusion (91.38% IRR; $\kappa=0.81$, 95% CI=[0.65, 0.97], indicating substantial agreement). Next, both authors conducted independent hand searches of any journal with two or more published studies included in our electronic search (i.e., *Journal of Behavioral Education*, *Journal of Positive Behavioral Interventions*, *School Psychology Review*) and found no additional articles for inclusion (IRR=100%; $\kappa=1.00$). Both authors then conducted independent ancestral searches of included studies’ references, yielding 43 titles to screen. We obtained abstracts, and of those, 13 studies were then obtained to read in full, with two additional articles identified for inclusion ($\kappa=0.94$, 95% CI=[0.91, 0.96], indicating near-perfect agreement) for a total of 37 included studies. Finally, we contacted corresponding authors and journal editors to inquire of any additional studies utilizing student-delivered BSP; while five articles were nominated from this step, no additional manuscripts were included. Later, when we began quality indicator coding, we realized one article (Wilson et al., 2001) involved counting tootles without ever sharing them aloud with students (thus students never heard peer praise intended for them), so we excluded it at that stage, resulting in 36 total studies.

Inclusion criteria

The included studies met six criteria. First, independent variable(s) included, primarily, student-delivered verbal or written BSP, defined as “providing students with praise statements that explicitly describe the behavior being praised” (Allday et al., 2012, p. 87), and was not packaged with other interventions (e.g., precorrection, Good Behavior Game, peer tutoring). Group contingency, self-monitoring, performance feedback, and other forms of increasing peer-delivered BSP were acceptable pseudo-packages (components of a peer-to-peer praise intervention). If students tootled to teachers about peers, studies were included when tootles were read aloud or somehow shared with peers later. Second, dependent variable(s) included at least one of the following student outcome measures: challenging behavior (e.g., disruptive behavior, problem behavior, aggression, off-task), time on task/academic engaged time, social skills, social interactions (including compliments and encouragements), and/or social status. Third, participants were school-age youth, general education or special education, from grades preK–12. Fourth, the intervention took place in a school setting, including university-sponsored laboratory schools (non-clinical) and alternative schools for students with severe behavior when part of a public or private school district.



Studies conducted in residential treatment centers were also included if the study took place in the school setting. Home settings, or clinics resembling classroom settings, were excluded as they were highly controlled settings, varying substantially from traditional school settings. Fifth, the study followed an experimental design: single case or group. Sixth, the study was a thesis, dissertation, or journal article available in English. We did not place a date restriction and accepted articles from any year.

Coding procedures

To understand both the rigor and relevance of the included studies, we conducted both quality indicator (QI) and descriptive coding. Both authors have published numerous quality assessment reviews, both together (e.g., Ennis et al., 2017; Royer et al., 2019) and separately (e.g., Royer et al., 2017; Ennis and Losinski, 2019); therefore, we did not code practice articles not included in this review prior to coding included studies. We met and reviewed the elements relevant to this review prior to coding, discussing potential nuances to QIs, then coded one article at a time independently

before meeting to discuss discrepancies and clarify QIs before coding the next study.

QI coding

We independently coded included articles for Council for Exceptional Children (2014) QIs of methodologically sound studies. QI 1.0 examines context and setting and we required studies to have at least one demographic variable to describe the setting that confirmed inclusion (e.g., school setting). QI 2.0 examines the participants and we again required at least one demographic variable for participants (2.1) and a description of why students or classes (depending on the case of analysis) were targeted for inclusion (2.2). QI 3.0 examines the intervention agent. Since student-delivered BSP typically involved implementation steps by adults and students, we required one demographic variable for each type of interventionist (3.1) and required evidence of both adult and student interventionist training, including an active check for understanding or use of a script to deliver the intervention (3.2). The remaining QIs did not require unique clarifications or distinction for this review, including (4.0) description of practice, (5.0) implementation fidelity, (6.0) internal validity, (7.0) outcome

measures/dependent variables (DVs), and (8.0) data analysis. Certain quality indicators are only applicable to either single-case (i.e., 6.5, 6.6, 6.7, 8.2) or group (i.e., 6.4, 6.8, 6.9, 7.6, 8.1, 8.3) design methodology and we applied them accordingly. For additional details on QI components, please see [Council for Exceptional Children \(2014\)](#).

We independently coded articles in a QI matrix ([Lane et al., 2019a](#)) in MS Excel one at a time, then compared and discussed any disagreements before coding the next. The mean IRR was 98.36% across all 36 articles (range = 89.29%–100%) and 97.11% by QI component (range = 80.00–100%). Overall κ for QI coding was 0.89 (95% CI = [0.83, 0.94]) indicating near-perfect agreement.

During QI coding, both authors independently made notes of descriptive characteristics of the studies that correspond to the [Council for Exceptional Children \(2014\)](#) QI. The first author's coding was used to create the descriptive table and the second author verified all information cell-by-cell, and while no errors were found, she suggested 24 refinements (out of 288 table cells) for easier readability. IRR for descriptive coding was 91.67%.

Evaluation procedures for classifying the evidence base of practices

For a study to be included in calculations for an evidence-based practice category, it had to meet 80% or more of QIs ([Lane et al., 2009](#)) using weighted coding, and if the study utilized single-case design, it had to include at least three cases (e.g., students, classrooms) and QI 6.5 had to be met (the design had to provide the *possibility* of at least three demonstrations of effect). We reviewed studies meeting these criteria and classified them as having either *positive*, *neutral* or *mixed*, or *negative* effects according to [Council for Exceptional Children \(2014\)](#) standards. For group studies, we used author-published effect sizes or calculated effect sizes when enough data were provided (e.g., n , M , and SD per group), then followed What Works Clearinghouse cut scores (as listed in [Council for Exceptional Children, 2014](#)) for *positive* ($d \geq 0.25$), *neutral* or *mixed* ($-0.25 < d < 0.25$), or *negative* ($d \leq -0.25$) effects.

We then used these classifications to determine if student-delivered BSP met [Council for Exceptional Children \(2014\)](#) criteria for an *evidence-based practice* (EBP), *potentially EBP*, *mixed evidence*, *insufficient evidence*, or *negative effects*. [Council for Exceptional Children \(2014\)](#) standards state an *evidence-based practice* (intervention, strategy, or practice scientifically validated through rigorous research methodology) has one of the following: (a) two group design studies utilizing randomized assignment with 60 or more participants, (b) four group design studies not utilizing randomized assignment with 120 or more participants, (c) five single-case studies (each with at least three participants and 75% or more showing therapeutic outcomes) with 20 or more total participants, or (d) a combination of group and single-case studies. Combinations can include one group randomized with 30 or more total participants and three single-case studies with 10 or more total participants, or two group non-randomized with 60 or more total participants and three single-case studies with 10 or more total participants. Additionally, no study can have *negative* effects and the ratio of studies with *positive* effects to *neutral* or *mixed* effects must be at least 3:1. More details about *potentially EBP*, *insufficient evidence*, and *negative effects*

category criteria can be found in [Council for Exceptional Children \(2014\)](#).

Data extraction and analysis

We calculated effect sizes for each dependent variable in group and single-case design studies that were eligible to contribute to the evidence-based practice classification (i.e., met our [Council for Exceptional Children, 2014](#) 80% weighted criterion, met QI 6.5, and had three or more cases if a single-case research design study). First, we extracted data from graphs using WebPlotDigitizer ([Rohatgi, 2024](#)) prior to performing analysis. When a study had multiple outcome measures, our primary focus was on outcomes of academic engagement/on-task behavior and disruptive behaviors. When study designs included multiple intervention conditions, such as students serving as peer praise recipient and peer praise teller (e.g., [Chenier, 2010](#)), we combined intervention conditions into one and compared those results to baseline.

For withdrawal/reversal and multiple baseline designs, we utilized a web-based calculator ([Pustejovsky et al., 2023](#)) to calculate between-case standard mean difference (BC-SMD) effect size estimates. For the one eligible alternating treatment design study ([Thoele, 2024](#)), we utilized a web-based calculator ([Manolov and Onghena, 2018](#)) to calculate an average difference between successive observations (ADISO) value. ADISO values can be standardized for comparison across studies by dividing by the standard deviation. For group design studies, we used author-provided n , M , and SD for each group to calculate Hedges's g . BC-SMD and standardized ADISO effect sizes are comparable to standardized mean differences from group comparison design studies ([Valentine et al., 2016](#)). Effect sizes were interpreted as small (0.20–0.50), medium (0.50–0.80), or large (≥ 0.80 ; [Fritz et al., 2012](#)). When determining if a single-case research design study had *positive*, *neutral* or *mixed*, or *negative* effects for consideration for the evidence base, we relied on the more conservative visual analysis in keeping with [Council for Exceptional Children \(2014\)](#) standards for evidence-based practices (as opposed to substituting our calculated effect size estimates).

We calculated both fixed-effect (assumes one true effect size underlies all studies; more weight given to larger studies with less variance) and random-effects (true effect size may vary across studies; studies with larger variances receive less weight) model ([Dettori et al., 2022](#)) meta-analyses for (a) all studies we were able to calculate an effect size for, (b) PPR studies separately, and (c) tootling studies separately, following formulas described by [Schluter \(2024\)](#). We constructed a forest plot of each study's dependent variables' effect sizes and the three overall peer praise category meta-analysis results following procedures demonstrated by [Lajeunesse \(2021\)](#).

Results

The 36 included studies represented 13 dissertations, four theses, and 19 journal articles, spanning from 1976 to 2024. The journal articles were published in 13 unique journals, with the *Journal of Behavioral Education* and the *Journal of Positive Behavior*

Interventions containing three articles each. Dissertations and theses represented 10 unique institutions, with University of Southern Mississippi and Louisiana State University each accounting for four dissertations/theses.

QI 1.0: Context and setting

All studies met QI 1.0 for context and setting by providing at least one detail about the school and/or classroom setting, allowing us to determine inclusion criteria (see Figure 2 for a summary of QI coding across studies). Published studies implemented peer-delivered BSP across the preK-12 continuum, with most taking place in elementary schools ($n=24$; see Table 1 for descriptive characteristics of all studies). Similarly, studies also took place across the least restrictive environment continuum, with most taking place in general education settings including whole school ($n=26$), followed by special education classrooms ($n=6$), and residential settings ($n=4$). Of note, many studies reported school- or facility-wide implementation of positive behavioral interventions and supports, with some even reporting school- or facility-wide fidelity scores (e.g., Sherman, 2012; Kennedy et al., 2014). While most studies took place in academic settings, a few studies took place in alternate settings, including the playground (Chenier, 2010; Teerlink et al., 2017) and homework time during after-school care (Kirkpatrick et al., 2019).

QI 2.0: Participants

All studies met QI 2.1 for providing at least one detail about study participants. 77.78% of studies met QI 2.2. for reporting details of why the student or class was targeted for intervention (e.g., disability status, challenging behavior, classroom management support needs). Many studies utilized data-based decision making to identify students for participation, with some studies confirming teacher or principal referrals of students or classrooms with direct observation screenings (e.g., Wright, 2019). Some studies utilized the class as a unit of analysis by pooling student data (e.g., Grieger et al., 1976), others examined the data of target students within classrooms (e.g., Ervin et al., 1996), and some studies reported both class and target student data (e.g., Lambert, 2014; McHugh et al., 2016).

QI 3.0: Intervention agent

For QI 3.1, 86.11% of studies met this QI by including demographics about both the adult and student (i.e., delivering BSP to peers) interventionists. However, only 63.89% of studies met QI 3.2 by providing sufficient information about the training of both interventionists. Some authors provided adults and/or students with a script to ensure fidelity of all implementation steps of the peer praise intervention—McHugh et al. (2016) even included procedures for rehearsing the script with feedback. Lum et al. (2019) is one example

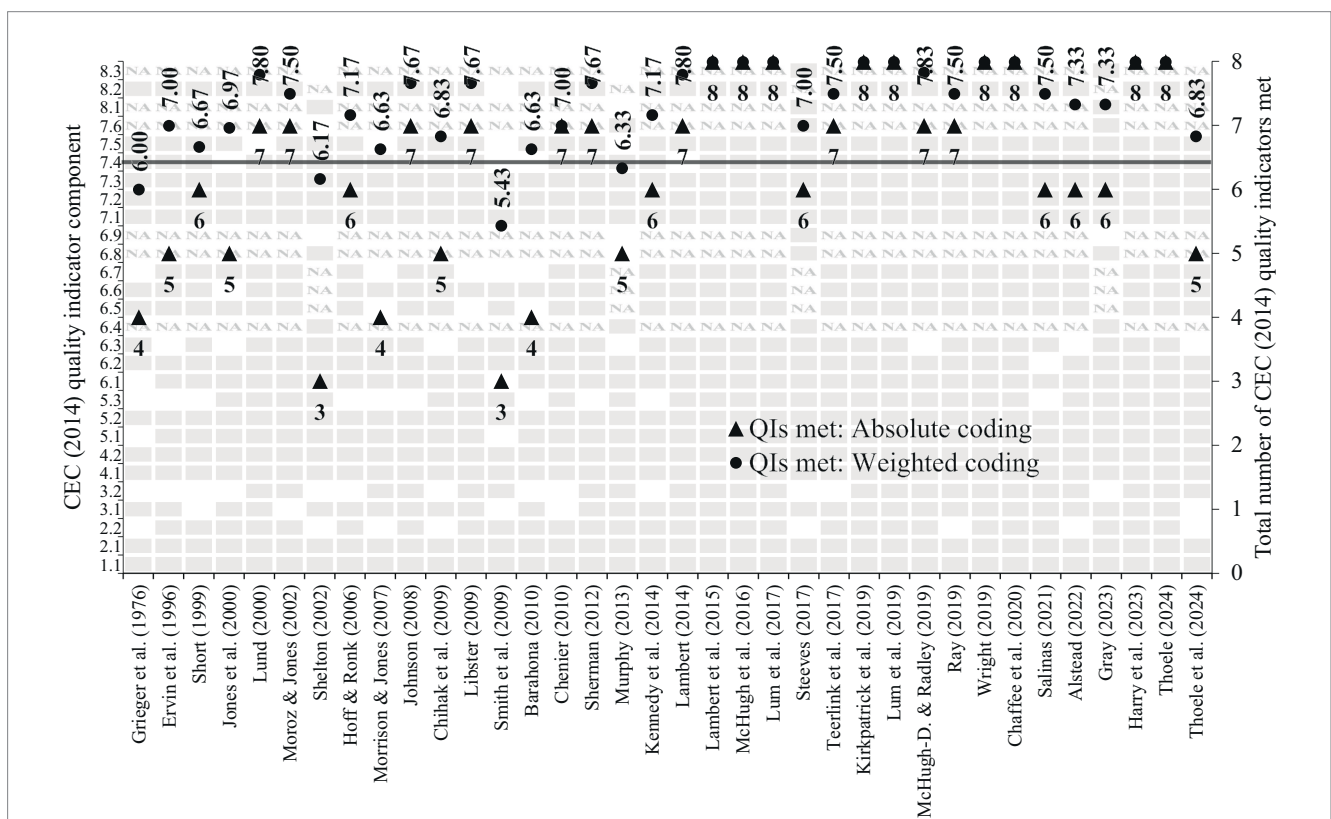


FIGURE 2 Methodological rigor of student-delivered behavior-specific praise (peer praise) studies. Peer praise studies are on the abscissa, and Council for Exceptional Children (2014) QIs met are on the primary ordinate (shaded cells=met, clear cells=not met). The secondary ordinate displays QIs met by absolute (triangles; 8.0 QIs required) and weighted (circles; 6.4 QIs required, 80%) coding to be considered methodologically sound. The weighted coding criterion of 6.4 is indicated by the horizontal black line. CEC=Council for Exceptional Children; QI=quality indicator.

TABLE 1 Descriptive characteristics and study effect classification (EC) for peer-delivered behavior-specific praise (peer praise) studies.

QI	Grieger et al. (1976)	Ervin et al. (1996)	Short (1999)	Jones et al. (2000)	Lund (2000)
1.0. Context	Two public school morning and afternoon K sessions	Middle school in the Family Home Program at Boys Town during math class	Home classroom middle school Boys Town with 8–11 children during fourth period before lunch and recess hour; obs in lunchroom	Midwest middle school residential program for pre- and delinquent adolescents; eighth grade math class with nine students and one teacher	Elementary school; four gened grade 5 classrooms where students rotated for reading, math, writing, and science
2.0. Participants	Ninety grade K students across four morning/afternoon blocks, 21–24 students each; selection criteria not described	Thirteen yo grade 7 student target, two others so not center of all attention; poor social skills and socially rejected by virtually all peers	Four female students 10–15 yo, three White, one American Indian; referred for peer problems, rejected by classmates, scored 3+ tattling Qs; teachers not described	Three 13 yo students, two Black male and one White female, referred for disruptive bx during cooperative learning activities	Eight grade 5 students (six male); nominated by teacher for out-of-seat and disruptive talking, verified by five 15-min obs
3.0. Intervention agent	Teacher/student praiser demo provided; training not described	Female teacher; student praiser demo provided; training not described	Teachers not described, training not described; student praiser demo provided; teachers told target students they would get points for PPR, gave examples; no CFU	Female math teacher 10 yrs. exp., training not described; student praiser demo provided, trained 20 min by teacher: rationale, steps, examples and non-examples	Interventionist = author; student praiser demo provided; author trained peer helpers privately: overview of procedures, modeling, role play with experimenter, and then each other until 80% accuracy for two consecutive training sessions
4.0. Description of practice	End of session opportunity to name a classmate (who received a happy face badge) and describe their friendly bx; no teacher praise during reporting	Teacher awarded students points, later exchanged for privileges, for positive, specific, direct, genuine comments about target students at end of math class for 5 min	Target student positively reported on by peers to teacher before recess/lunch, then teacher privately shared the BSP with complimented peer and recorded a point on point card	3 days/week during 30-min cooperative learning math worksheet activities (groups of 3–4), students received points for praising specific bx of “star” student at end of class (5–7 min), and the star could praise others; steps for PPR on bulletin board	Peer helpers (1/2) praised engagement (in-seat quiet) paired with a token every 3 min; token earners (1/2) exchanged daily for tangibles or activities; generalization without tokens, then roles reversed in a different classroom, then different classroom with no procedures; prize options = small tangible (e.g., pencils, stickers) or short activity (e.g., extra computer time, board games)
5.0. Imp. fidelity	No fidelity reported	Researcher checklist 37% of intervention sessions = 100% fidelity	Researcher checklist for one baseline and at least one intervention session per student = 100% fidelity; one fidelity check was the first intervention session	Researcher checked peer compliments contained three or four of the four steps, and all peers present provided compliments with two exceptions; confirmed teacher awarded points each day	Researcher observed peer helper for integrity of appropriate token delivery and praise = 97% accuracy
6.0. Internal validity	A-B-A'-B' reversal design (A' report unfriendly children; B' no happy face badge available); baseline = free play	A-B-A-B reversal design; baseline = math class	MBL across participants; baseline = standard Boys Town operating procedures, which included points to a student who reported a peer's appropriate bx	Non-concurrent MBL, baseline was immediately before turn as star; baseline = class divided into three groups with instructions to cooperatively complete multiskill assignments 3× week	MBL across participants; baseline = each teacher taught reading, math, writing, or science, and all students rotated between them

(Continued)

TABLE 1 (Continued)

QI	Grieger et al. (1976)	Ervin et al. (1996)	Short (1999)	Jones et al. (2000)	Lund (2000)
7.0. Outcome measures/DVs	SV = no measure; DV = aggressive acts frequency and cooperative play 1-min MTS (median % of students playing with each other) recorded on 23 days during separate 15-min sessions; IOA = 16 observers trained to reliability (0.90); reliability checks repeated at 2-week intervals indicated the observers maintained this level during all phases of the study	SV = no measure; DV = 9–10 min direct obs 15-s partial interval recording for positive or negative peer interactions during math; IOA 35% obs = 96% positive, 99% negative interactions	SV = TEI-SF teacher post; DV = sociometric questionnaire (author-created) student pre post follow-up; peer nomination (author-created) pre post follow-up; CBCL TRF and YSR teacher pre; AIR-PS student pre; peer report frequency; 15-min direct obs 15-s partial interval recording for positive, negative, neutral lunch interactions; Boys Town point card (in/appropriate bx frequency); IOA 30% obs = 87%+	SV = IRP-15 teacher post; sociometric rating (Oden, 1980) of likelihood of spending free time with each peer; DV = 30-min direct obs 15-s interval frequency recording cooperative statements; IOA 45% obs = 85%	SV = researcher created 5 Q student post, 10 Q pre- and 16 Q teacher post; DV = 10–15 min direct obs 10 s observe 5 s record interval system for teacher attention, peer attention, engagement (>1/2 an interval), passive off-task (>1/2 an interval), out-of-seat, inappropriate verbal, aggression, token delivery, verbal praise; IOA 27% sessions = means >80%, ranges 50–100%
8.0. Data analysis	Cooperative classroom plays increased from 42 to 55%, then from 42 to 60%; aggressive acts decreased from 42 to 9 and then from 40 to 6 in each A-B sequence	Baseline = high negative interactions and low positive interactions; intervention = near zero negative interactions and positive interactions >70% of intervals	Positive interactions did not increase, no changes to negative interactions; no change for social status on SRS-PNF	Cooperative statements and peer status increased for all students	Engagement increased dramatically, disruptive bx decreased, and somewhat generalized to other classrooms
EC	n/a, 80% of QI not met	n/a, only one case	Neutral or mixed effects	n/a, only two cases	Positive effects
QI	Moroz and Jones (2002)	Shelton (2002)	Hoff and Rank (2006)	Morrison and Jones (2007)	Johnson (2008)
1.0. Context	Public Midwest elementary school, three classrooms	Urban elementary school in southeast, 80% FRPL, mostly White, 20% Black	Public elementary grade 3 and 4 sped classroom in a medium-sized Midwest community; one teacher, one aide, and seven students; desks in large cluster with students facing each other	Inner-city Title I elementary school; 95% FRPL	Suburb of major metro city southeastern USA, grade 2 gened classroom with 20 students, predominantly Black
2.0. Participants	Three White female students aged 7 (grade 1 gened), 8 (grade 3 gened), and 10 (grade 3 sped ED SLD); referred by teacher for severe social withdrawal, isolation, low peer interaction	Intervention: 12 female and 8 male grade 3, 8 female and 6 male grade 4; control: 9 female and 9 male grade 3, 8 female and 7 male grade 4; assistant principal referred teachers who might be interested	Three male, four female grade 3 and 4 students with ID (IQ 50–75); teacher self-nominated class because students struggled to get along	Two grade 3 gened classrooms with 13 students (12 Black, 1 White; 1 ID) and 14 students (13 Black, 1 White; 1 ED); selection criteria not described	Fifteen grade 2 students aged 7–8; teacher selected for various student bx issues she wanted to address
3.0. Intervention agent	Teacher demo not provided, description of PPR provided at screening, given script; student praiser demo provided, trained by teacher using script, provided examples to show understanding	Four female grade 3 and 4 teachers, instructed to record number of tootles, place check next to date after announcing tootles, and provided description of what constitutes tootles with examples to share with class, but no CFU; student praiser demo provided, trained by experimenter 20+ min, provided examples of tootles to show understanding	Female classroom sped teacher, training not described; student praiser demo provided, trained by author for 40 min twice over 2 weeks: compliment delivery, PPR steps, memorize a compliment, practice, create poster, timed how quickly they could generate specific genuine compliments	Two female gened teachers, followed PPR script; student praiser demo provided, trained by teacher and research team for 30 min: rationale, description, steps, students say examples and non-examples, all students gave 1 correctly	Author (female school counselor) implemented PPR; female teacher's role was fidelity so trained how to do that; student praiser demo provided, trained by researcher with explanations, descriptions of prosocial bxs for two 20-min sessions, and students gave examples, received feedback

(Continued)

TABLE 1 (Continued)

QI	Maroz and Jones (2002)	Shelton (2002)	Hoff and Ronk (2006)	Morrison and Jones (2007)	Johnson (2008)
4.0. Description of practice	Star student praised by peers for specific good bx 7–10 min between morning recess and academics; teacher praised appropriate PPR rewarded with sticker on chart (sped) or cotton ball toward popcorn group contingency (gened)	Index card taped to student desks for tootles, turned in end of day or when full; researcher announced total next day, shared examples, praised and corrected, reviewed process; experimenter emailed teachers daily count to announce in morning and update cardboard goal ladder; class reward = 15 min extra recess, popsicles	Teacher selected MVP name from bag start of day, displayed on poster, given plush toy and token box on desk; students watched for good bx and social interactions and complimented, placing token in box; last 10 min of day students reported MVP good bxs, teacher praised; tokens totaled weekly, 10 filled a square (120 total) on pyramid of success to earn cupcake party; students created poster of compliment sentence starters	15 min daily before lunch, teacher used script, passed out numbered notecards with steps for BSP on back, reviewed steps, students made examples and non-examples; teacher spun wheel and student with number drew chance card with directive to praise student to right, student the teacher selects, or receive praise from student with higher number; candy or sticker delivered to both students; chance cards read to remaining students in order until lunch	2 h for students to obs peers, then 10 min to report peer prosocial bx; researcher praised appropriate reporting, praiser and praised received sticker
5.0. Imp. fidelity	Teacher daily checklist = 100% fidelity, and count of peer BSP (as a measure of fidelity) averaged 17, 21, and 9 per day each classroom	Teacher daily checklist = 100% fidelity for counting tootles, announcing tootles to class, and updating goal ladder	Researcher checklist 46% of intervention condition obs (vs. end of day reporting) = 93% fidelity	Researcher script completion check 52% of sessions = 99% fidelity	Teacher checklist 29% of intervention sessions = 100% fidelity
6.0. Internal validity	MBL across participants with a reversal; baseline = teachers handled interactions between students in usual manner	Pretest–posttest comparison group design, one classroom per grade randomly selected to receive the tootling intervention	A-B-A-B withdrawal design; prior to the study, author led 16 40-min weekly skillstreaming social skills group meetings, but teacher reported minimal maintenance of social skills; baseline = 30-min unstructured morning free time with choice of activities: socializing with peers or aide, journal writing, games, coloring, seat work, or reading	MBL across two classrooms; baseline = routine classroom bx management plan of posted rules, teacher prompts and warnings, negative consequences	A-B-A-B withdrawal design; baseline = math question and answer session, modified at intervention start to be centers where students could interact and work in small groups
7.0. Outcome measures/DVs	SV = IRP-15 teacher post; DV = 30-min direct obs 10-s partial interval recording for social involvement during recess; teacher perceptions of social withdrawal and isolation on ASCA; IOA 31% obs = 92%, $\kappa = 0.87$	SV = no measure; DV = prosocial/antisocial attention and recognition measure (researcher created) where students watched 10-min video of students, then 10 min to describe what they saw, scoring specific bxs mentioned; peer perception scale (researcher created); 20% of responses on each scored by second rater, IRA = 95% and 100%	SV = no measure; DV = 28-min direct obs 15-s partial interval recording rotating students each 30 s for prosocial and negative social interactions during morning free time; IOA 18% obs = 98%, $\kappa = 0.78$	SV = no measure; DV = teacher-recorded low- and high-intensity bxs on adapted CEI, avg. daily score per week used to monitor each class and each lunch and two transition periods; sociometric nominations: students named three children they would like to play with, one or less nominations = rejected/neglected; IOA 15% of lunchtime generalization obs = 91%	SV = researcher created teacher parent pre-post, student post with smiley neutral negative faces; DV = 11–20 min direct obs 8-s whole interval recording rotating students for prosocial interactions; IOA 24% obs = 94%, 97%, 90%, 94% by condition

(Continued)

TABLE 1 (Continued)

QI	Moroz and Jones (2002)	Shelton (2002)	Hoff and Ronk (2006)	Morrison and Jones (2007)	Johnson (2008)
8.0. Data analysis	All students increased social engagement during intervention; only two students returned to baseline levels once withdrawn	Repeated measures MANOVAs; no increased awareness of peer prosocial bx, no increase in positive perception of classmates, students more aware of peer antisocial bx	Prosocial peer interactions increased from 16% to 26%, then 17% to 24% each A-B sequence; negative interactions were low and stable throughout conditions 1–2%	CEI low-frequency high-intensity bxs reduced in both classrooms (avg. 4.17–3.17 per day and 10.72–7.87 per day); 1 class experienced ~2 fewer critical events during lunch, the other ~4 fewer after treatment	Prosocial interactions increased from 16% to 59%, then from 37% to 45% in each A-B sequence
EC	Neutral or mixed effects	n/a, 80% of QI not met	n/a, only one case	n/a, QI 6.5 not met	n/a, only one case
QI	Cihak et al. (2009)	Libster (2009)	Smith et al. (2009)	Barahona (2010)	Chenier (2010)
1.0. Context	Rural elementary Title I school in southeast, grade 3 inclusive classroom	Grades 1 and 2 gened public elementary classrooms in Baton Rouge, Louisiana 25 students each classroom	Three gened preschool Head Start classrooms in the south with 20 students 4–5 yo	Three grade 3 gened classrooms in Louisiana	Public and private schools in east Baton Rouge Parish grades 1–4
2.0. Participants	Eight female, 11 male, grade 3 students, 4 with SLD and/or ADHD, 3 Black, 14 White, 2 Latinx	One Black male grade 1 (shy, no participation, recess alone), one Black female grade 2 (active, aggressive, recess alone); whole class; participants selected for being neglected; teachers volunteered after researcher met with grade 1–3 teachers	One student from three classrooms: White (5 yo female), White (4 yo male), Black (4 yo female); peer sociometric ratings = lowest or second lowest in class, peer rejected	Grade 3 classroom; teacher or principal referred for disruptive classroom bx	Grade 1 into grade 2 male in private school, grade 4 female and grade 2 male in public school; SSI-S-RS scores provided; teachers nominated socially withdrawn students, confirmed by peer sociometric ratings and playground obs of 80%+ alone time
3.0. Intervention agent	Certified elementary teacher 3 yrs. exp., training not described; student praiser demo provided, trained by teacher for two 20-min sessions how to tootle, wrote examples, and received praise or feedback until all were successful	Female teacher, training not reported, teacher asked researcher (female doctoral student) to take over PPR end-of-day sessions; student praiser demo provided, trained by researcher and teacher for 15 min: how to praise specifically, modeled examples and non-examples, students listed examples for pro-social bxs	No teacher demo, trained at monthly staff meeting on procedures, modeling, had questions answered, and used fidelity checklist as script when implementing; student praiser demo provided, trained by teacher using a script, students practiced praise statements	Female classroom teachers, study purpose and teacher responsibilities explained by researcher (no training described); student praiser demo provided, trained by researcher and teacher for 15-min session per intervention, learned to correctly report and write prosocial praise statements, received praise and feedback until accurate examples from all students were generated	No teacher demo, training not described; student demo provided for targeted participants, peer praisers only had grade level reported, trained by teachers: described rules, told to observe star during day report end of day, example positive bx, no CFU
4.0. Description of practice	Teacher reviewed tootling procedures start of day, placed 4" × 6" index cards on desks, encouraged tootling, collected cards prior to each transition; 20 min before end of class teacher read tootles aloud (number not reported), counted and updated poster goal; group contingency goal met = extra recess time	Researcher explained class would PPR to earn points toward pizza party by raising hand and praising the star at the end of each day; poster of praise examples and reinforcement chart tracking PPR points	Daily 15-min PPR sessions students earned a token each time someone voluntarily gave a positive statement to the star student of the week; tokens filled jar to earn popsicle party	PPR days teacher read student examples, praised, pulled three stars, names on board, end of day 10 min students praised prosocial bx of stars; if 10, earned token for extra recess time; tootle days students wrote index cards at end of day for three stars for 10 min, teacher counted (no read aloud), gave to stars; after 1 week, two classes started unknown stars; posters of example praise and "what is praise?"	Each student (worker bee) praised a peer (king/queen bee for 1 or 2 weeks) end of day <10 min, earned a token (pollen) for the container (beehive); reports were public and teacher reinforced praised bx; 30 tokens earned small reward, 130 earned pizza party; rules posted

(Continued)

TABLE 1 (Continued)

QI	Cihak et al. (2009)	Libster (2009)	Smith et al. (2009)	Barahona (2010)	Chenier (2010)
5.0. Imp. fidelity	Researcher checklist 40% of school days during each phase = 99% fidelity	Researcher checklist 33% of PPR sessions = 94% fidelity	Teacher daily checklist for intervention condition, results not reported	Researcher checklist 22–43% of intervention days, depending on classroom: fidelity for PPR = 90%, 98%, 92% fidelity and tootling = 80%, 82%, 100% fidelity	Teacher and researcher checklist daily = 100% fidelity
6.0. Internal validity	A-B-A-B withdrawal design; baseline = typical grade 3 classroom	Non-concurrent MBL across two subjects; baseline = grade 1 and 2 typical classroom and recess	MBL across participants with embedded A-B-A and follow-up probe; baseline = typical preschool Head Start classrooms	Alternating treatments design with initial baseline and control condition, random selection for sequence; two classrooms did not have five alternations; baseline = classroom instruction, classroom activities, small group activities with teacher current classroom management plan during	Non-concurrent MBL design with baseline, praise recipient (king/queen bee), and teller (worker bee) conditions; baseline = typical elementary playground activities
7.0. Outcome measures/DVs	SV = IRP-15 teacher post; DV = all day teacher direct obs frequency recording disruptive bx on construction paper bracelet listing student initials; IOA 30%+ obs each condition = 92% for all students; IRR 30% of permanent product scoring of tootles meeting criteria = 100%	SV = no measure; DV = direct obs class activity and recess (obs length not reported) 15-s partial interval recording for social interaction; peer acceptance (social status and sociometric rankings via peer nomination and ratings); IOA 40% obs = 95% classroom, 89% recess	SV = IRP-15 teacher post (moderate results, large variability); DV = social acceptance (students placed pictures of peers in smiley, neutral, and frowning face boxes based on how much they liked playing together); 15-min direct obs frequency recording for negative bxs; IOA not reported	SV = IRP-15 teacher post, CIRP student post each condition; DV = 10- to 13.5-min direct obs 10-s partial interval recording rotating students for disruptive and on-task bx; direct bx rating scales: followed directions and on task 90% of day, positively interacted; IOA 21–30% obs = 93%, 93%, 95% per class	SV = IRP-15 teacher post, CIRP student post; DV = pre post sociometric rating scale (Coie et al., 1982) for participants by peers; pre post SSIS-RS; direct obs on playground (obs length not reported) 15-s MTS for positive, negative, and neutral social interactions; IOA 51% obs = 96%, 94%, and 94% per student
8.0. Data analysis	Disruptive classroom bx decreased from 23.2 to 8.4, then from 16 to 3.5 each A-B sequence with 0 disruptive bxs last 3 days	Both students improved quality of social interactions (female decreased negative classroom interactions from 40% to <1% of intervals, male increased positive classroom interactions from 8% to 69%) and peer acceptance	One of three students increased social acceptance based on picture sociometric ratings by peers, moving from least in class to second most preferred; other two students slightly decreased in social acceptance; two of three students reduced negative bxs and maintained after PPR, other student was already low in baseline	Neither tootles nor PPR (both known or unknown stars) had a significant impact on disruptive classroom bx	Average increased % of intervals in positive social interactions = 26%; SSIS-RS increased M = 11 for social skills and M = 16 for top 10; sociometric ratings changed for one participant who was no longer rated as least liked by any classmates
EC	n/a, only one case	n/a, QI 6.5 not met	n/a, 80% of QI not met	n/a, QI 6.5 not met	Neutral or mixed effects
QI	Sherman (2012)	Murphy (2013)	Kennedy et al. (2014)	Lambert (2014)	Lambert et al. (2015)
1.0. Context	Rural southeast school in year 4 of PBIS with overall SET 98.2; gened elementary classrooms grades 3–6 with 15–22 students	Alternative public school (Continuous Learning Center) grades 6–12 in Mobile, Alabama with 89 students with sped needs removed from neighborhood school due to problem bx 45+ days, 71% male, 78% Black, 92% FRPL	Residential facility grades 1–12 with high-fidelity PBIS; three art classes with one sped teacher and two bx specialists	Two schools in rural southeast with high-fidelity PBIS, one with two grade 6 classrooms and one school's grade 7 classroom; 19–28 students per classroom, gened or inclusion	Two elementary schools in southeast with high-fidelity PBIS, grade 4 and 5 gened classrooms with 17–19 students, one majority Black one majority White

(Continued)

TABLE 1 (Continued)

QI	Sherman (2012)	Murphy (2013)	Kennedy et al. (2014)	Lambert (2014)	Lambert et al. (2015)
2.0. Participants	Four gened grade 3–6 students: 9 yo Black female, 9 yo Black male, 11 yo White female, 8 yo Black female; principal referred students for inappropriate classroom bx and social difficulties, screening obs confirmed	Sixteen students, 13–16 yo, grade 7 or 8 math, grade 8 language arts, 6–7 students per class; classes selected in collaboration with school administrators based on consistency, fit, and teacher willingness	Eight elementary students 7–11 yo EBD grades 2–4; disruptive bx function = positive reinforcement: attention, 5+ ODR, direct obs confirmation; art class selected for frequency of inappropriate bxs	Eleven yo Black female, 12 yo Black female, 13 yo Black male; administrator and teacher referred classrooms, teachers nominated students for disruptive bx confirmed by screening obs	Grade 5 gened classroom 19 students, 9 female, 10 male, 15 White, 1 Asian, 1 Latinx, 2 Black; grade 4 gened classroom 17 students, 10 male, 7 female, 2 SLD, 13 Black, 3 White, 1 Latinx; administrators referred classrooms for disruptive bx confirmed by obs
3.0. Intervention agent	Four female teachers with 1–23 yrs. of exp.; teachers trained with scripts, practiced implementing, and provided feedback; student praiser demo provided, trained by teachers for 30-min following a script, students gave examples that were praised or given feedback	One Black male, two White female teachers 3–7 yrs. exp., trained by experimenter using session fidelity checklist for one to two 30-min sessions with description of PPR, rationale, steps, handouts, practice with students until all gave a correct praise; student praiser demo provided, trained by teachers with experimenter help: modeling, sample praise statements, students gave examples, discussed when to praise	Sped art teacher 3 yrs. exp., 2 bx specialists, training not described; student praiser demo provided, trained by teacher: appropriate peer relations, praise notes, modeled, practiced with teacher and peer feedback	Three Black female teachers with 3 or 9 yrs. exp., two master's degrees, trained by researcher, given scripted procedures; student praiser demo provided; trained by teachers using script: how to observe and tootle appropriate bx, examples and non-examples, practice with feedback or praise until all tootled correctly	Two White female teachers, one with master's degree 9 yrs. exp. and other with bachelor's degree 1 year exp., trained by researchers, given scripted procedures; student praiser demo provided, trained by teachers using a script: recognize and report peer appropriate bx, gave examples and non-examples, students practiced and received praise or feedback until each student tootled correctly
4.0. Description of practice	Students observed star (announced start of day) all day for appropriate bx with hourly reminders; each PPR for star (~10 min) earned token to fill jar; PPR + tootling: students could also write tootles for star throughout the day and place in shoebox, read by teacher end of PPR time and praised along with token; tokens earned party, movie, extra recess; poster board of PPR steps	Compared business as usual, teacher PPR, researcher PPR; individual praisers earned stickers, cotton ball by praising 2 weekly MVPs rotated alphabetically last 5 min each day; 280 cotton balls = pizza party; appropriate praise posters, small dry-erase board with MVP name and score, plastic achievement medals	After art instruction, teacher wrote praise notes (star shape) to students or students had 10–15 min to write PPN (every student received at least one); students read their praise notes before peer activity time (board games, cards, puzzles, art)	Teacher gave students index cards start of day with instructions and encouragement to tootle (one on front, one on back, more cards if needed) appropriate peer bx throughout class period, some randomly read by teacher end of period, all added to count toward class goal (chips, ice cream sandwiches, donuts, game day, extra recess time)	Teacher gave index cards to students at start of 2-h period, reviewed instructions, and encouraged tootling (one on front, one on back; one classroom could get extra cards) of peer-appropriate bx, read by teacher (5+) end of period, teacher praised, added to group reward (extra recess, 20 min computer time, cupcakes, chips)
5.0. Imp. fidelity	Researcher checklist ~30% of intervention sessions = 88–100% fidelity per teacher per condition; 30% of observed sessions had fidelity IOA = 100%	Researcher checklist 3+ sessions per week and results graphed to compare the two conditions: high experimenter involvement (78.3% fidelity) vs. low (<2% fidelity)	Researcher checklist 43–46% of sessions = 95%, 96%, 99% fidelity per art class, IOA 50–61% of observed sessions = 90%, 100%, 93%	Teacher daily checklist = 95%, 94%, 95% fidelity; observer checklist during obs = 96%, 94%, 97% fidelity, IOA 40–44% of obs = 100%	Teacher daily checklist = 100%, 97% fidelity; researcher checklist each obs during intervention = 97%, 100% fidelity; IOA for researcher 37–49% obs per classroom = 100%; training fidelity checklist = 100%
6.0. Internal validity	MBL designs across two sets of two participants counterbalanced for order effects; baseline = teachers dealt with inappropriate bx normally in typical elementary classrooms	Quasi-experimental group design; obs Friday Game Time, unstructured, casual; business as usual group teacher followed routine classroom bx management plans: posted rules, prompts, and warnings, negative consequences; some students changed classrooms in the middle of intervention	Alternating treatments design comparing PPN and teacher praise notes; obs last 15 min of art class during peer activity time to practice social and play skills (board games, cards, puzzles, art)	A-B-A-B withdrawal design; baseline = teachers continued normal classroom routines and bx management techniques during math, language arts, world history classes	A-B-A-B withdrawal designs with a MBL element across two classrooms; baseline = teachers continued normal instruction and classroom management procedures across multiple subjects

(Continued)

TABLE 1 (Continued)

QI	Sherman (2012)	Murphy (2013)	Kennedy et al. (2014)	Lambert (2014)	Lambert et al. (2015)
7.0. Outcome measures/DVs	SV = IRP-15 teacher post, students asked what they thought about each intervention; DV = CBS subscales for aggression, prosocial, asocial, anxiety, peer rejection, hyperactivity; 20-min direct obs 10-s partial interval recording for inappropriate and appropriate bx; IOA 30%+ of obs = 82–95% across students and bxs	SV = IRP-15 teacher pre post, CIRP student pre post; DV = author created sociometric rating of peers pre post; 25-min direct obs 15-s partial interval recording for positive, negative, neutral, or no social interactions during unstructured game time each Friday, observed at individual student level but analyzed at classroom level; IOA not reported	SV = author created 4–5 question teacher students unit-supervisor post surveys for each student; DV = 15-min direct obs duration of inappropriate bxs during art classes; IOA 30–38% obs = 96% or 100% per student	SV = IRP-15 teacher post, CIRP students post; DV = 20-min direct obs 10-s MTS for appropriate and disruptive bx, target student observed every third interval; IOA 25–50% obs = 91%, 88%, 93% per class	SV = IRP-15 teacher post; DV = 20-min direct obs 10-s MTS for appropriate and disruptive bx in science or language arts; IOA 33–60% each condition = 91%, 93% per class
8.0. Data analysis	No differences between PPR and PPR + tootling, both reduced inappropriate bx for all four students and increased appropriate bx for three students	Positive interactions increased 34% high researcher involvement class, decreased 14% low researcher involvement class, decreased 22% control class; sociometric ratings increased for all three classes by 11%, 5%, 13%, respectively	Teacher praise notes decreased inappropriate bxs an average 34% and PPN 36%	Overall positive effects for all three target students; reduced classwide disruptive bx and increased classwide appropriate bx	Disruptive bx decreased and appropriate bx increased substantially for both classrooms during tootling phases, with moderate to strong non-overlap of all pairs
EC	Positive effects	n/a, 80% of QI not met	Positive effects	Positive effects	n/a, only two cases
QI	McHugh et al. (2016)	Lum et al. (2017)	Steeves (2017)	Teerlink et al. (2017)	
1.0. Context	One grade 2 and two grade 3 gened elementary classrooms in two southeast schools implementing high-fidelity PBIS; classes = 9–11 male, 9–12 female, 11–19 Black, 0–11 White, 0–3 Latinx, 0–3 ELL OHI	Rural southeast high school, 590 students, 68% FRPL, three gened classes on four 95-min block schedule	Two public elementary schools in Baton Rouge, Louisiana, one in a low-income area of the city, one for students with reading-related disabilities; grade 3 and 4 classrooms	Suburban elementary Title 1 school western USA; fourth year of SWPBS; 55% male, 83% FRPL, 38% ELL, 32% sped; 52% Latinx, 39% White, 4% Pacific Islander, 2% Black, 1% American Indian, 1% Multiple	
2.0. Participants	8 yo Black female; 7 yo Black male, 8 yo Black male; principals referred disruptive bx, teachers nominated a most disruptive target student; classrooms screened confirmed 30%+ intervals of disruptive bx	Student sex, grade, race, sped provided in table each classroom; school administrators referred classrooms for disruptive bx, confirmed by 30%+ intervals of disruptive bx in screening obs	113 students grade 3 (31%) grade 4 (69%), 43.4% female, 67.3% Black, 21.2% White, 8% Latinx, 1.8% Asian, 1.8% other, avg. 9.87 yo (range = 8–12 yo); school selection criteria not described; teachers recruited on interest and willingness	167 peer praisers (42% grade 4, 36% grade 5, 22% grade 6); seven White female recess aides and one Latinx male administrator provided recess supervision; whole school 32.38% of all ODR = playground; 1–2 teachers each grade 4–6 chose 2–3 students who behaved well and 2–3 students with challenging bx	
3.0. Intervention agent	White female teachers 1 or 8 yrs. exp., one master's degree, trained by researcher on components, given script to train students, rehearsed, asked Qs, received feedback; student praiser demo provided, trained by teachers using script: monitor and write peer appropriate bx, examples, practice, feedback until all made one correct tootle	Three White teachers 22–30 yo 1–2 year exp., female English literature or physical science, male geometry, trained by researcher with script to train students, read together, answered Qs; student praiser demo provided, trained by teachers using script: observe and record peer prosocial bx, examples and non-examples, practice writing, read aloud for class with feedback	Nine grade 3 or 4 gened teachers, 88.9% female, 66.7% White, 22.2% Black, 11.1% Latinx, 10.78 avg. yrs. exp., 11.1% master's degrees, training stated but not described; student demo provided for all groups combined, training not described beyond introduced students to the procedures, no CFU	School administrator (author) male Latinx; student praiser demo provided, trained by author, and supervised in distributing PPNs during recess: 1-h interactive PowerPoint reviewed playground rules, role and responsibilities, procedures, schedule, mystery motivator, time to practice, no CFU	

(Continued)

TABLE 1 (Continued)

QI	McHugh et al. (2016)	Lum et al. (2017)	Steeves (2017)	Teerlink et al. (2017)
4.0. Description of practice	Teachers passed out note cards during the most problematic time period (20, 30, or 60 min) for students to tootle appropriate peer bxs, reminded students of procedures; students turned in tootles to a box and got more cards, working toward low daily goal (25–30 tootles) reachable if each wrote 2 per day; end of tootling teacher updated number on thermometer, read at least 5 to class, praised appropriate bx; class reward = extra recess time, show and tell, small edibles and tangibles	Teacher passed out tootle paper at period start, instructed and encouraged students to tootle appropriate peer bx all period one per slip, put in container, get new slip; teacher read 5+ tootles end of period, added to count toward goal (homework passes, movie, cookies, chips, donuts); observer performance feedback on missing steps	Classrooms randomly assigned control, gratitude, tootling; gratitude = students made three gratitude statements each morning for the previous day at school after teacher modeled, and teacher wrote praise notes end of day so each student got 1+ per week; tootling = teacher reviewed tootling data each morning from day before, read sample tootles, praised, opportunities to practice tootling, reviewed class progress, then students kept note cards on desks all day and looked for positive peer bx, turned in tootles end of day	12–15 peer praisers at recess looked for responsible, respectful, safe students, verbally stated bx observed, completed PPN for student
5.0. Imp. fidelity	Teacher daily checklist = 100%, 100%, 98% integrity; researcher checklist, frequency not reported = 100%, 93%, 90% fidelity, with IOA 33–50% of sessions = 100%; training fidelity checklist for training each teacher = 100%	Teacher daily checklist = 82%, 100%, 90% fidelity; researcher implementation checklist day 1 for each teacher = 100% fidelity; researcher fidelity across obs = 81%, 100%, 98%, IOA for 33%	Teacher daily checklist = 86% gratitude fidelity 95% tootling fidelity with 2 days (13% intervention sessions) researcher completed IOA = 100%; 485, 143, and 63 tootles per class; gratitude teachers wrote note to each student once per 5 days	PPN triplicate copies served as fidelity, checking praisers distributed expected number of PPN evenly across assigned grades = 35% of expected PPN distributed
6.0. Internal validity	A-B-A-B withdrawal with a MBL element across two classrooms; baseline = classroom teachers used schedules, routines, and bx expectations; obs during language arts (silent reading, responding to questions about readings as a class, independent worksheets, quizzes) or science (brief lesson, hands-on activity, worksheet)	A-B-A-B withdrawal design with follow-up 1–2 weeks later; baseline = teachers managed classrooms in their typical manner, routines, bx management techniques	quasi-experimental repeated measures group design; classrooms randomly assigned to one of three conditions: control, gratitude, tootling	A-B-A-B withdrawal design; baseline = typical elementary recess
7.0. Outcome measures/DVs	SV = IRP-15 teacher post, CIRP target student post; DV = 20-min direct obs 10-s partial interval recording for disruptive and academically engaged bx, target student observed every third interval; IOA 25–66%, obs = 92–94% across classes	SV = BIRS teacher post; DV = 20-min direct obs 10-s MTS rotating students for disruptive and academically engaged bx; IOA 42% = 89%, 91%, 94% per class	SV = URP-IR teacher pre post, CIRP student pre post; DV = STRS-SF pre post follow-up; BBRS pre post follow-up; SSWQ student pre post follow-up; SEHS-P student pre post follow-up; ODR; weekly classroom conduct grades; alpha and/or reliability coefficients provided for measures, construct validity discussed	SV = researcher created survey recess aides school administrator students (15 each grade randomly selected) post; DV = playground ODR; PPN count (triplicate copy to office, teacher, and student); IRR not reported
8.0. Data analysis	Disruptive bx decreased and academically engaged bx increased in all three classrooms and target students	Disruptive bx decreased and academically engaged bx increased across classrooms	Mixed-design two-way MANOVAs: no significant main effect for SSWQ; significant main effect for SEHS-P: gratitude decreased in control and persistence increased for all conditions; significant main effect for STRS-SF: closeness increased for control and gratitude, conflict increased for control; mixed repeated-measures ANOVA for BBRS = significant main effect: decrease for control and significant improvement for both tootling and gratitude; no substantial change in conduct grades; ODRs decreased all conditions, gratitude most from 14 to 0 vs. tootling 5 to 1 vs. control 8 to 1	Median ODR per day decreased by 42.9%; when PPNs were reintroduced after withdrawal, median ODR per day decreased to 0
EC	Positive effects	Positive effects	Neutral or mixed effects	n/a, only one case

(Continued)

TABLE 1 (Continued)

QI	Kirkpatrick et al. (2019)	Lum et al. (2019)	McHugh Dillon and Radley (2019)	Ray (2019)
1.0. Context	University-assisted urban after-school program in the southern USA, K-5, 49% female, 51% male, 42% Black, 41% White, 15% Latinx, >60% FRPL, 3% ELL, 18% sped	Rural southeast high school on 95-min block schedule, 550 students, 68% FRPL	Two rural elementary schools in southeast	Rural K-8 school in upper Midwest
2.0. Participants	Grade 3 after-school program classroom, four Black (two female, two male) students 8–10 yo, three extra reading instruction; female teachers not certified; program director referred grade 3 for mean and disrespectful bx toward peers and staff; teacher nominated students aggressive, disrespectful, mean bx, confirmed by researcher informal obs	Algebra II, accelerated English II, English IV; 13–17 female, 7–12 male, grades 10–12, 15–16 White, 5–11 Black, 0–1 Latinx, 0–5 sped SLD OHI; school administrator referred gened classrooms with disruptive bx, confirmed by screening obs 30%+ intervals classwide disruptive bx	Three elementary classrooms obs in language arts, science, or math: 8–20 female and 9–15 male students, 0–24 Black, 10–19 White, 0–1 multiracial, 0–5 sped SLD OHI ADHD 1 ASD; principals referred gened classrooms with disruptive bx, confirmed by screening obs 30%+ of intervals with disruptive bx	Three sped classrooms, 10 students grades 2–5 with IEP bx goal: 0–4 male, 0–1 female, 0–4 Latinx, 0–3 White, 0–1 Black, EBD OHI SLD ASD ID; teachers invited, chose to participate
3.0. Intervention agent	Four female researcher interventionists, 23–30 yo, 2–3 yrs. PhD school psychology program; student praiser demo provided, trained by researchers for 40 min: tootling as a game to earn extra recess, defined tootling, examples and non-examples, how to write tootles, students gave examples and received feedback	Three White female 23–30 yo teachers 1–6 yrs. exp., trained by researcher: reviewed script to use with students, examples and non-examples, practice with feedback until accurate; student praiser demo provided by class, trained by teachers using script, wrote tootles, received feedback, modeled submitting	Three White, one Black female teachers 1–20 yrs. exp., one specialist's, one master's degree, trained by researcher for 15-min: explained tootling, modeled ClassDojo, script for training students, rehearsed, feedback; student praiser demo provided, trained by teachers for 15 min using script, examples and non-examples, practiced saying tootle while entering ClassDojo, feedback until all correct	Three White female teachers 1–19 yrs. exp., one White male counselor 4 yrs. exp., researcher trained using student training script; student praiser demo provided, trained by teachers for 20 min using script, examples and non-examples, students distinguished tootles from tattles, practiced with vignettes and modeling, students wrote tootle with feedback
4.0. Description of practice	Researchers reviewed tootling start of academic hour, read 1–2 previous day examples, praised class if met criterion and placed a star on reward board (three stars earned extra recess); student drew criterion from bag for how many different students needed to receive a tootle that day; 3" × 5" index cards provided, placed tootles in box during academic hour, last 10 min of class students practiced writing tootles while researchers provided feedback	Teacher placed 12 cm × 6 cm slips on desks, encouraged students to write peer prosocial bx observed during class, submit (tear in half: student praiser name in one container, peer praised and bx in other container) and get new slip; five rewards given out daily end of class when teacher picked three tootles to read and pulled two praiser names; rewards = homework passes, candy bars, chips, soda	Teacher encouraged students to notice peer appropriate bx throughout the 20-min session and enter tootles on ClassDojo computer stations end of session (5 s per student; 3 min total max) which updated interactive whiteboard with names and tootles; teacher praised progress toward goal (free time, extra recess time, popcorn, skittles)	Teacher put 4" × 5.5" blank tootles by container, reviewed instructions, encouraged students to tootle throughout the day if they saw peer prosocial bx; start of each morning social skills group, teacher read all tootles aloud, updated goal thermometer (cupcakes, popcorn, freeze pops, cookie party, pizza party, extra recess time, touch football)
5.0. Imp. fidelity	Researcher checklist 40% of each condition = 100% intervention fidelity, 0% baseline and withdrawal fidelity	Teacher daily intervention checklist = 100% fidelity; observer checklist every obs = 100% fidelity, 42% with IOA = 100%	Researcher daily checklist = 96%, 95%, 99% fidelity, IOA for 33% = 100%; teacher daily checklist = 100%, 100%, 97% fidelity; teachers training students fidelity = 100%	Teacher daily checklist = 100% fidelity; researcher checklist on obs days = 94% fidelity; teacher training students observed by researcher with checklist = 98% fidelity, with IOA for 1 classroom = 100%; authors noted students wrote few tootles

(Continued)

TABLE 1 (Continued)

QI	Kirkpatrick et al. (2019)	Lum et al. (2019)	McHugh Dillon and Radley (2019)	Ray (2019)
6.0. Internal validity	A-B-A-B withdrawal design; baseline = 20 min on carpet reviewing upcoming activities and receiving teacher-led instruction, then went to stations for spelling, reading, math with rules relaxed because of afterschool program, so punishment applied less consistently	A-B-A-B withdrawal design; baseline = teachers continued typical classroom routines and bx management strategies in Algebra II, accelerated English II, English IV	A-B-A-B withdrawal design; baseline = teachers conducted typical classroom management techniques to handle disruptive and appropriate bx in language arts, science, math	MBL design across classroom settings; baseline = social skills period in the morning
7.0. Outcome measures/DVs	SV = no measure; DV = 20-min direct obs 15-s partial interval recording one target student at a time in sequence for antisocial and disrespectful bx during carpet time academic hour; IOA 33–60% per condition = 90–98% per condition	SV = BIRS teacher post, CIRP students post; DV = 20-min direct obs 10-s MTS for disruptive, academic engaged, and passive off-task bx; IOA 40% = 94%, $\kappa = 0.79$	SV = BIRS teacher post; DV = 20-min direct obs 10-s MTS for disruptive and academic engaged bx; IOA 40% obs = 96%, 94%, 94% per class, $\kappa = 0.93, 0.87, 0.88$	SV = IRP-15 teacher post; DV = frequency count of appropriate tootles; 30-min direct obs 10-s MTS for on-task bx, 10-s partial interval recording for disruptive and prosocial bx; IOA 30%+ of obs = 99% each class
8.0. Data analysis	Tootling decreased antisocial and disrespectful bx with Tau-U = 0.84 and 0.93	Tootling resulted in immediate decreases of disruptive bx and increases in academic engaged bx across all three classrooms during intervention; passive off-task bx remained relatively stable and low across conditions	Tootling on ClassDojo resulted in immediate decreases of disruptive bx and increases in academic engaged bx during intervention in all three classrooms	Low number of tootles each classroom ($k = 4, 5, 6$; only two classes met goal of five tootles, once); tootling increased on-task bx (high to begin with) and decreased disruptive bx (low to begin with); prosocial bx increased for one classroom
EC	Positive effects	Positive effects	Positive effects	Positive effects
QI	Wright (2019)	Chaffee et al. (2020)	Salinas (2021)	Alstead (2022)
1.0. Context	Rural high school in southeast on 90-min block schedule with ~600 students, 65% White, 31% Black, 49% female, 51% male, 70% FRPL	Middle school in a northeast metropolitan area on 42- to 49-min block schedule, 614 students, 9.6% FRPL, 6.5% ELL, 19.1% sped	Three public school self-contained sped classrooms for students with behavioral concerns, 4–6 male and 0–3 female students per classroom	Middle school in upper Midwest, inclusive social studies class last hour of day
2.0. Participants	Three gened geometry or English classrooms grades 9–12: 4–7 female, 14–16 male, 13–17 White, 3–5 Black, 0–2 Latinx, 0–1 American Indian, 3–8 sped SLD OHI VI; teachers requested classroom management support, confirmed by 20-min screening obs <70% on-task bx	Two grade 6 ELA and inclusion social studies classrooms: 17–24 students, 54–59% male, 59–67% White, 21–35% Asian, 6–12% Black, one to two 504 plan ADHD, 0–4 sped OHI ADHD TBI; teachers contacted school psychologist disruptive bx support, confirmed by 20-min screening obs 30%+ of intervals disruptive bx	Classrooms with grades 1–5, 6–8, 9–12, two male ED students targeted in each: 9–16 yo, 1 Black, 1 White, 4 Latinx, IQ60–88; selection criteria not described	34 students, 5 sped, grade 5, social studies class; principal referred teacher excited to participate and suggested social studies would have social interaction during group work; teacher expressed concerns for on-task bx, academic engagement, disruptive bx, and following schoolwide expectations

(Continued)

TABLE 1 (Continued)

QI	Wright (2019)	Chaffee et al. (2020)	Salinas (2021)	Alstead (2022)
3.0. Intervention agent	Two White male and female teachers 2–3 yrs. exp., one master's degree, trained by researcher for 20-min using student training script and opportunity for role play; classroom peer demo provided, trained by teachers using script, determined rewards, nominated 2–3 student interventionists per class trained by researcher: provided daily checklist	Two White male and female teachers 31–54 yo master's degrees, 6–30 yrs. exp., trained by researcher: tootling introduction, modeling, role play, script for training students, script for daily implementation; student praiser demo provided, trained by teachers who described tootling as competition, students practiced writing tootles with feedback	Two female and one male teacher 37–39 yo 3–6 yrs. exp. one master's degree, trained by researcher: examples and non-examples, script for training students; student praiser demo provided, trained by teachers: description of tootling, bxs to tootle, Padlet, practice, goal and rewards, students submitted a valid tootle	Female teacher 30+ yrs. exp., first-year social studies, trained by researcher for 30 min: described tootling, how implemented, how to train students, answered Qs, role play before each new phase; student praiser demo provided, trained by researcher and teacher for 30 min with script: examples and non-examples, procedures, all students wrote a proper tootle, received feedback
4.0. Description of practice	Student interventionist placed two tootle slips on desks, announced and posted goal, placed envelopes in front of class with chance slips and mystery motivator, encouraged student to observe peer positive bx during 90-min class block, place tootles in container; end of class read 5 and added total to goal; if daily goal met drew chance slip (2:1 ratio of reward to X), praised meeting goal, encouraged for tomorrow, opened mystery motivator if reward chance slip was drawn; rewards = chips, candy, bonus points, free time	Teachers gave students 3" × 5" index cards and encouraged tootling of peer appropriate bx; end of class 5 min collected tootles, read 5 aloud, praised students receiving the tootle, updated goal thermometer; rewards = 15-min recess, choice of seating	Teacher told students to use their devices to submit tootles via Padlet any time all day; end of day teacher displayed Padlet tootles, counted, updated Google Sheets bar graph; laminated chart of tootling bx examples, rewards = ice cream, chips, candy bars, \$5 fast food gift card, pizza	Start of class only students get tootle cards (1/3 of 8.5" × 11"), tootle prosocial bx, turned in last 5 min of class; reward = ice cream sundae party, pizza party, donut party, cupcake party
5.0. Imp. fidelity	Student interventionist daily checklist = 81%+ fidelity; researcher daily checklist = 100% fidelity, IOA 30% sessions = 100%; all trainings = 100% fidelity and IOA: researcher training teacher, teacher training classroom, researcher training student interventionists	Teacher daily checklist completed irregularly, results not reported; observer daily checklist = 96%, 93% intervention 100% baseline withdrawal fidelity per class, IOA 38–48% sessions = 100%; researcher training teacher and teacher training students = 100% fidelity	Teacher daily checklist, results not reported; researcher observed 40% of sessions = 100% fidelity; researcher observed teachers training students = 100% fidelity	Researcher checklist once per phase (20%+ per condition) = 100% fidelity; same checklist guided teacher implementation
6.0. Internal validity	A-B-A-B withdrawal; baseline = teachers followed normal classroom routine and bx management strategies for geometry (beginning of each block) and English (middle of block)	A-B-A-B-C reversal design with maintenance; baseline = typical instruction and classroom management practices, including flexible seating, logical consequences, verbal prompting, loss of lunch/recess privileges in ELA or social studies	A-B-A-B withdrawal design; baseline = teachers continued instruction as usual and tended to events as standard	A-AB-A-AB-ABC-AB-ABC-ABCD-ABC-ABCD reversal design: (A) wrote tootles (B) teacher reported number of tootles each morning (C) public posting of goal thermometer (D) teacher read three tootles start of class with feedback and praise for praiser and recipient; no baseline, social studies group assignments, individual projects, videos, whole-class instruction

(Continued)

TABLE 1 (Continued)

QI	Wright (2019)	Chaffee et al. (2020)	Salinas (2021)	Alstead (2022)
7.0. Outcome measures/DVs	SV = BIRS teacher post, CIRP student post; DV = 20-min direct obs 10-s MTS for academically engaged and disruptive bxs; IOA 34–41% obs = 97%, 97%, 96% per class	SV = URP-IR teacher post-maintenance, CURP student post-maintenance; DV = 20-min direct obs 15-s partial interval for disruptive bx, 15-s MTS for academic engaged bx; IOA 38–46% obs = 92%, 93% per class	SV = BIRS teacher post, research created questionnaire student post; DV = 20-min direct obs 10-s partial (disruptive bx) or whole (academically engaged bx) interval recording; IOA 40% for each condition = 95–98% across classes	SV = IRP-15 teacher post, students asked by teacher about feelings on acceptability, what they liked/did not like/would change; DV = 15-min direct obs 10-s MTS for on-task bx and 10-s partial interval recording for disruptive bx; IOA 30% obs = 98.6% on-task bx 100% disruptive bx
8.0. Data analysis	Two of the three classrooms increased academic engaged bx and decreased disruptive bx	Academic engaged bx improved in class A, results unclear in class B due to paraeducator staffing change; disruptive bx reduced when tootling was introduced but stayed low throughout remaining conditions	Immediate therapeutic changes in disruptive bx and academic engaged bx across all three classrooms	On-task and disruptive bx improved from beginning to end of study; mean on-task bx was 67.39% of intervals at start, >88% of intervals on average by end; disruptive bx had decreasing trend, was low to begin, and stayed low near zero
EC	neutral or mixed effects	n/a, only two cases	positive effects	n/a, only one case

QI	Gray (2023)	Harry et al. (2023)	Thoele (2024)	Thoele (2024)
1.0. Context	Charter K-8 southeastern urban school in a very large and diverse district (233,000+ students, refugee, majority Latinx or Black); 448 students, 98.9% FRPL, 52.9% male, 42.2% Black, 36.8% Latinx, 13.8% White, 1.8% Asian, 5.1% multiracial; PBIS in place	Rural southeast high school, 600 students, 68% FRPL, four gened classrooms	Rural southeast elementary school, 551 students; three grade 4 or five classrooms during ELA or science (one cotaught inclusion, one sped): 8–11 White, 5–9 Black, 4–5 Latinx, 0–4 multiracial, 0–3 ELL, 4–10 sped	Rural southeast elementary school, self-contained sped classroom
2.0. Participants	26 students in grades 5–8 (14 control, 12 intervention), 50% Black, 30.8% Latinx, 54% male; college faculty recommended charter schools and first principal accepted offer; students screened for self-reported life satisfaction and teacher bx screener, plus administration and teacher referral	Algebra, health, biology, English classrooms: 4–13 female, 5–15 male, 8–10 White, 4–9 Black, 0–2 Latinx, 0–1 Asian, 0–1 Pacific Islander, grades 9–11, 0–5 sped SLD OHI ASD SLI; teachers self-referred classroom management and bx problems, confirmed by screening obs <70% intervals on-task bx	One or two students targeted per class: 10–11 yo Black or multiracial males, two not sped, three sped OHI ADHD ODD SLD; principal referred classrooms for classroom management and bx concerns; teachers nominated target students with or at risk for EBD, confirmed with screening obs 30%+ intervals disruptive bx	Five students in classroom with bx goals on active IEPs, three assented: two 10 yo Black one 8 yo White; school or classroom selection criteria not described
3.0. Intervention agent	Two graduate students facilitated seven small groups, no demographics, trained on PPR but no CFU, received weekly guidance and support; student praiser demo provided, trained using PPR protocol script, no CFU	Three White male, one female teacher 1–7 yrs. exp. one master's degree, trained by researchers for 30 min following script, modeled steps, answered Qs, two teachers rehearsed; student praiser demo provided, trained by teachers for 20 min: procedures, examples and non-examples, practiced writing with feedback, voted on rewards	Three White female teachers 4–19 yrs. exp. master's or education specialist degree; class demo provided; interventionist groups trained by researcher with slides on prosocial bx, rules, goal, modeling, two students and one teacher practiced script with feedback, all students wrote tootle, five reviewed by teacher with feedback, voted on rewards	White female teacher 15 yrs. exp. master's degree; consented students demo provided but not other peers; both intervention groups trained by researcher for 30 min with slides on prosocial bxs, tootling rules, goal for rewards, modeling, two students and teacher practiced with feedback

(Continued)

TABLE 1 (Continued)

QI	Gray (2023)	Harry et al. (2023)	Thoele (2024)	Thoele (2024)
4.0. Description of practice	Students randomly assigned to culturally adapted Well-Being Promotion Program (Suldo, 2016) or the program with integrated PPR; PPR group reminded at each session to look for positive bxs aligned with PBIS expectations and character strengths of two students chosen as stars, names on whiteboard; end of session students reported positive bxs of stars, each PPR counted toward goal (10 PPR per student)	Teacher gave two tootle papers at class start, encouraged students to report two daily in container; end of day teacher silently read tootles, added to goal poster; BC phase teacher posted after each day on bulletin board who and what bxs were tootled; tootle example poster	Teacher- or student-led tootling; leader read script reminding students to look for and tootle prosocial bx; tootling partner assigned to each student so duo would tootle on each other, new partners each week; end of session leader read three tootles, praised recipient of the tootle and praiser; when goal met, leader spun choice wheel and class immediately received reward; teacher placed tootles in student folders to go home to parents after goal was met; students voted with colored squares every fourth session for no tootling, teacher-led, or student-led; rewards = pencils, chips, extra recess, candy, computer time, wear a hat pass	Students and teacher randomly selected to lead tootling sessions: read tootling script to remind students to look for and tootle prosocial bxs to earn class goal, pass out 3" × 5" tootle slips with reminder of components, collect and count tootles to place in envelope; if goal of 10 tootles met, leader spun choice spinner and immediately delivered class reward; students who wrote and received tootles received BSP from teacher or researcher privately at end of each session; choice condition = students and teachers voted on student-led tootling, teacher-led tootling, or no intervention; rewards = candy, pens, computer pass, wear hat in class, sit on couch, extra recess time
5.0. Imp. fidelity	Researcher checklist per session = 99% fidelity	Teacher daily checklist = 100% fidelity; researcher checklist each obs = 100%, 100%, 97%, 95% fidelity by class, with IOA 25%+ obs but not reported; teacher training of students = 88–100% fidelity, with 100% IOA	Researcher checklist for 81–88% of intervention sessions = 90–100% fidelity, with IOA for 36–38% of sessions = 100%; researcher checked tootles = 95% met criteria	Researcher checklist 98% of sessions = 95% fidelity
6.0. Internal validity	Pretest–posttest group comparison design; stratified (grade level) random assignment by group generator	MBL design across classrooms with embedded A-B-BC (baseline, tootling, tootling + public posting); baseline = teachers followed normal classroom routine and bx management procedures in algebra, English, biology, health	Alternating treatments design with maintenance; conditions randomized in systematic blocks = baseline, student-led tootling, teacher-led tootling, student choice; baseline = teacher delivered usual instruction and bx management systems	Alternating treatments design within an A-B-A-C design (baseline, teacher-led and student-led tootling alternating, withdrawal, student choice); baseline = teacher delivered typical instruction and bx management techniques, including individual bx point sheets for student IEP bx goals (response cost)
7.0. Outcome measures/DVs	SV = CURP student post; DV = pre post student completed: MSLSS friends domain for peer relationship satisfaction, SLSS, PANAS-C, PROMIS for internalizing anxiety and depression, and SDQ hyperactivity conduct problems peer problems subscales	SV = BIRS teacher post, CIRP student post; DV = 20-min direct obs 10-s MTS for academically engaged, disruptive, and passive off-task bx; IOA 33%+ of obs = 97%, 98%, 97%, 94% per class	SV = researcher created tootling efficacy questionnaire teacher pre post, TAST teacher post, researcher created treatment acceptability questionnaire student post; DV = 20-min direct obs 10-s MTS for disruptive, passive off-task, and academic engagement bx; IOA 32–50% of sessions = 81–94% across students	SV = students and teacher voted for student-led tootling, teacher-led tootling, or no tootling each choice day; DV = 20-min direct obs 10-s MTS for disruptive and academic engagement bx; IOA 38% of sessions per condition = 89% disruptive bx, 93% academic engagement

(Continued)

TABLE 1 (Continued)

QI	Gray (2023)	Harry et al. (2023)	Thorsle (2024)	Thorsle (2024)
8.0. Data analysis	Mixed-model ANOVA—life satisfaction: no significant main effects; positive and negative affect: no significant main effects; externalizing bx: no significant main effects; anxiety and depression: no significant main effects	Academic engagement increased, disruptive bx decreased, and passive off-task bx decreased for all four classrooms; no differences between tootling and tootling with public posting	Teacher- and student-led tootling decreased percentage of intervals disruptive bx; both increased academic engagement but considerable variability; no meaningful difference for three students, one student responded better to teacher-led tootling, one student responded better to student-led tootling	Student-led tootling was as effective as teacher-led tootling in decreasing disruptive bx and increasing academic engagement
EC	Neutral or mixed effects	Positive effects	Positive effects	n/a, only one case

ADHD, attention deficit hyperactivity disorder; ASD, autism spectrum disorder; avg, average; BSP, behavior-specific praise; bx, behavior; CFU, check for understanding; demo, demographics; DV, dependent variable; EBD, emotional and behavioral disorder; ED, emotional disturbance; ELA, English language arts; ELL, English language learner; exp, experience; FRPL, free or reduced-price lunch; gened, general education; ID, intellectual disability; IOA, interobserver agreement; IRA, interrater agreement; K, kindergarten; MBL, multiple baseline; MTS, momentary time sampling; MVP, most valuable person; n/a, not applicable; obs, observation; ODR, office discipline referral; OHI, other health impaired; PBIS, positive behavioral interventions and supports; PPN, peer praise note; PPR, positive peer reporting; QI, quality indicator; Qs, questions; SET, Schoolwide Evaluation Tool; SLD, specific learning disability; SLI, speech language impairment; sped, special education; SV, social validity; SWPBS, school-wide positive behavior supports; VI, visual impairment; yo, years old; yrs, years. AIR-PS, Assessment of Interpersonal Relationship—Peer Scale (Bracken, 1993); ASCA, Adjustment Scales for Children and Adolescents (McDermott et al., 1993); BBRS, Brief Behavior Rating Scale (Gresham et al., 2010); BIRS, Behavior Intervention Rating Scale (Von Brock and Elliott, 1987); CBCL, Child Behavior Checklist (Achenbach, 1991); CBS, Child Behavior Scale (Ladd and Proffitt, 1996); CEI, Critical Events Index (Walker and Severson, 1992); CIRP, Children's Intervention Rating Profile (Coi et al., 1982; Witt and Elliott, 1985); CURP, Children's Usage Rating Profile (Britesch and Chatouleas, 2009); IRP-15, Intervention Rating Profile-15 (Martens et al., 1985); MSLSS, Multidimensional Student Life Satisfaction Scale (Oden, 1980; Huebner and Gilman, 2002); PANAS-C, Positive and Negative Affect Scales for Children (Ebesutani et al., 2012); PROMIS, Patient-Reported Outcomes Measurement Information System (Irwin et al., 2010); SDQ, Strengths and Difficulties Questionnaire (Goodman, 2001); SEHS-P, Social Emotional Health Survey—Primary (formerly titled the Positive Experiences at School Scale [PEASS]); Furlong et al., 2013); SLSS, Student Life Satisfaction Scale (Huebner, 1991); SSIS-RS, Social Skills Improvement System Rating Scales (Gresham and Elliott, 2008); SSWQ, Student Subjective Wellbeing Questionnaire (Renshaw et al., 2015); STRS-SF, Student-Teacher Relationship Scale—Short Form (Pianta, 2001; Suldo, 2016); TAST, Treatment Acceptability Survey for Teachers (Tandi et al., 2010); TEI-SF, Treatment Evaluation Inventory—Short Form (Kelley et al., 1989); TRF, Teacher Report Form (Achenbach, 1991); URP-IR, Usage Rating Profile—Intervention Revised (Chafouleas et al., 2011); YSR, Youth Self-Report (Achenbach, 1991).

of many where authors assessed and reported fidelity of the training steps to ensure researchers remembered to execute all training steps with all interventionists. A few authors, including Wright (2019), even reported training integrity with IOA for researchers training teachers and teachers training students.

QI 4.0: Description of practice

100% of studies met QI 4.1 and 4.2 by including adequate details on study procedures and materials. Within the 36 studies examining student-delivered BSP, there was some variation among intervention procedures. Twenty studies examined the tootling intervention, 15 examined positive peer reporting, two studies compared the two approaches (Barahona, 2010; Sherman, 2012), and three studies evaluated peer praise outside of positive peer reporting or tootling procedures by using written peer praise notes or peer helpers' verbal praise (Lund, 2000; Kennedy et al., 2014; Teerlink et al., 2017).

QI 5.0: Implementation fidelity

For QI 5.1, an impressive 94.44% of studies assessed and reported implementation fidelity data. 100% of studies met QI 5.2 for either directly reporting dosage or reporting information from which dosage could be inferred (e.g., graphed data with estimated daily dosage). However, only 72.22% of studies met QI 5.3, as some studies did not include language making it clear that fidelity was assessed throughout the intervention and/or intervention phases. Of note, Steeves (2017) utilized exemplary procedures for tracking dosage in a group design study, having teachers self-report daily implementation fidelity outcomes. Lambert (2014) and Lambert et al. (2015) both collected IOA of implementation fidelity data between two observers to ensure accuracy, a robust procedure though not required by Council for Exceptional Children (2014) QIs.

QI 6.0: Internal validity

QIs 6.1, 6.2, and 6.3 refer to both group and single-case research design studies. QI 6.1, met by 94.44% of studies, refers to the researcher's ability to control the independent variable. As an exemplar, Lum et al. (2017) included procedures during withdrawal phase for explicitly telling teachers to remove all intervention materials (e.g., tootle submitting container, poster of group contingency progress) and tell students the class was not going to tootle, if asked. QI 6.2, met by 97.22% of studies, refers to adequate description of baseline/comparison conditions. Both Hoff and Ronk (2006) and McHugh et al. (2016) provided detailed descriptions of not only how data were collected during baseline conditions but also what instructional procedures occurred (e.g., weekly social skills meeting; science brief lessons with hand-on activities and worksheets). QI 6.3, met by 80.56% of studies, refers to baseline/control conditions having no or extremely limited access to the independent variable. We marked QI 6.3 as not present in studies that did not include explicit mention of removing materials, telling teachers not to implement, and/or limiting access to the intervention in control/withdrawal conditions.

An exemplar, Kirkpatrick et al. (2019), included assessing fidelity of baseline and withdrawal conditions to report that 0% of implementation steps were implemented.

Within the 36 included studies, there were four (11.11%) that utilized group research design methodology. Of those four, 100% met QI 6.4 for clearly describing/utilizing best practices for group assignment, 75.0% met QI 6.8 for reporting (or allowing our calculation of) overall attrition, but only 25.0% (i.e., Steeves, 2017) met QI 6.9 for reporting directly or including enough data to allow us to calculate differential attrition.

Of the 32 (88.89%) single-case research design studies, 29 (90.63%) met QI 6.5 for using an experimental design that provided for the possibility of at least three demonstrations of effect. 31 studies (96.88%) met QI 6.6 for including at least three data points in all baseline conditions, and 27 studies (84.38%) met QI 6.7 for utilizing a design that controls for common threats to internal validity.

QI 7.0: Outcome measures/DVs

100% of studies met QIs 7.1, 7.2, and 7.3 for utilizing socially important outcomes, clearly defining dependent variables and their measurement, and reporting effects of all dependent variables. 88.89% of studies met QIs 7.4 and 7.5 for utilizing appropriate timing of dependent variable data collection (i.e., group designs close to end of intervention, single-case three or more data points per condition) and providing adequate evidence of group measure reliability or IOA of single-case research design direct observation dependent variables. Of note, McHugh et al. (2016) and Lum et al. (2017) impressively reported κ to account for chance agreement between two raters. QI 7.6 refers to group design methodology, and 75.0% of the four included studies met this QI for including adequate evidence of validity. For example, both Murphy (2013) and Gray (2023) included measures of social validity, as did Steeves (2017), who additionally discussed construct validity.

QI 8.0: Data analysis

QIs 8.1 and 8.3 apply to group design methodology. Of the four group studies in this sample, 75% met QI 8.1 for employing appropriate data analysis techniques, and 50% met QI 8.3 for reporting measures of effect or sufficient information from which we could calculate effect sizes. QI 8.2 applies to single-case research design methodology and requires studies to include a clear graph reporting data from all conditions for each unit of analysis. Of the 32 included single-case studies, 96.88% met this QI.

Evidence base supporting student-delivered behavior-specific praise

Based on Council for Exceptional Children (2014) standards for EBPs, tootling met criteria two times for classification as an *evidence-based practice* by having a minimum of five single-case research design studies with 20+ participants and also by having at least one group design study with 30+ participants and at least three single-case research design studies with 10+ participants. PPR did not meet

criteria for *evidence-based practice*, potentially EBP, nor *mixed evidence* because only one of the five single-case research design studies that met Council for Exceptional Children (2014) weighted criteria for methodological rigor had *positive* effects while four were *neutral* or *mixed* effects—two studies with *positive* effects were needed for the *mixed evidence* category, and so we classified PPR into the *insufficient evidence* category.

Figure 3 contains a forest plot of estimated effect sizes for all studies meeting 80% or more of QIs, our weighted Council for Exceptional Children (2014) criterion for methodologically sound studies. Each study is marked by the type of student-delivered BSP intervention employed: tootling ($k=20$), positive peer reporting ($k=15$), and other ($k=3$), with two studies marked as both PPR and tootling given authors compared the two interventions. The forest plot concludes with overall effect sizes for student-delivered BSP, inclusive of all studies and categories of peer praise, and then we considered the evidence base for PPR and tootling separately given the large and clear divide of studies into these categories. The important work of the three studies utilizing direct peer praise was inadequate in number for consideration of a separate evidence-based practice categorization or omnibus effect size.

Discussion

It was encouraging to find 34 of the 36 studies (94.44%) met QI 5.1 for reporting implementation fidelity results, as some past systematic literature reviews found very few studies met this important QI (e.g., 47.92% of studies coaching educators to increase BSP in Ennis et al. (2020); 46.15% of instructional choice studies in Royer et al., 2017). Results across the studies included in this systematic literature review showed student-delivered BSP can improve academic engaged time and reduce the disruptive behavior and social isolation of students with or at risk for EBD. Even PPR studies, which had *mixed evidence* in terms of Council for Exceptional Children (2014) standards for EBP, showed most individual student participants improved on multiple outcomes; there were just not enough participants in studies (minimum needed is three), studies did not meet QI 6.5 (study design provides for the possibility of at least three demonstrations of intervention effect), or <75% of participants showed improvement (see Table 1), and thus those studies could not be considered in the EBP classification calculations. Individual students who were socially withdrawn/rejected increased their social interactions when they received peer-delivered BSP as the star in PPR studies (e.g., Short, 1999; Chenier, 2010). Such results showed how student-delivered BSP can help increase the number of classroom positive interactions and support teachers who might not always be able to give as much attention to quiet students as they would like, perhaps because they feel drained of energy from going “through the same cycle with the [disruptive] behavior kids” over and over (Lanza, 2020, p. 36).

It was interesting to note the clear shift in studies from PPR to tootling in 2014, though it is unclear why the shift occurred at that time. Skinner et al. (1998) introduced the concept and name tootling at a 1998 conference, and the first tootling study was by Shelton (2002) a few years later. The next tootling study was Cihak et al. (2009), then researchers compared tootling to PPR in Barahona (2010) and Sherman (2012), with the final PPR study a year later by Murphy (2013) and all others

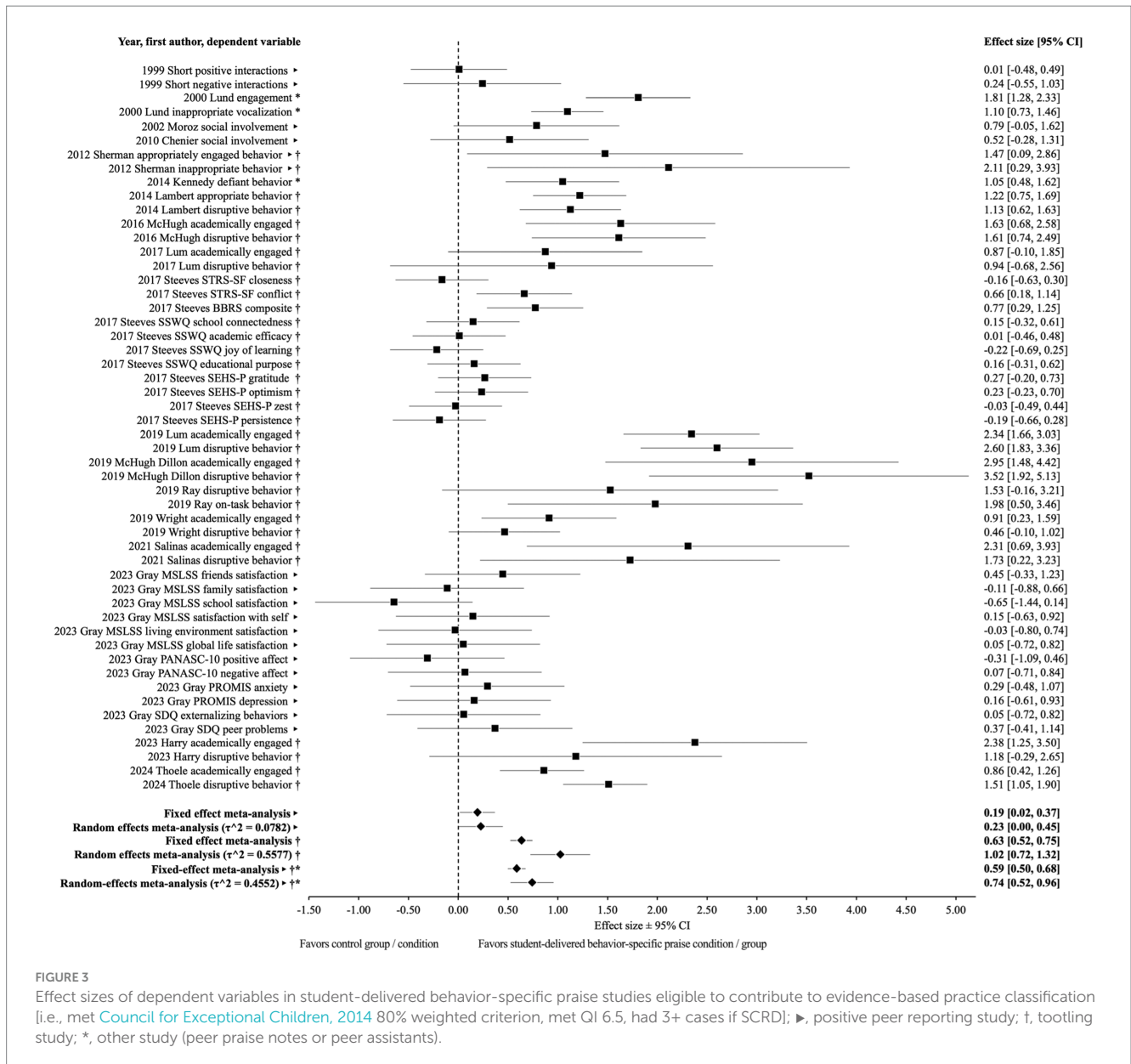


FIGURE 3 Effect sizes of dependent variables in student-delivered behavior-specific praise studies eligible to contribute to evidence-based practice classification [i.e., met Council for Exceptional Children, 2014 80% weighted criterion, met QI 6.5, had 3+ cases if SCRDI]; ▶, positive peer reporting study; †, tootling study; *, other study (peer praise notes or peer assistants).

through 2024 being tootling except an outlier 2023 dissertation (Gray, 2023). Perhaps this follows an ‘evolution’ in student-delivered BSP: from a single student being the PPR ‘star’ receiving all peer BSP, to having three students as stars, to having the star(s) be unknown so more students engage in expected behaviors hoping peers will notice and report on them later if they end up being the star(s), to scaling up peer praise classwide with tootling where all students are now observed by peers. We understand why teachers might prefer tootling because all students can receive BSP from their peers. In PPR studies, all students were reminded of behavior expectations when they were told to be on the lookout for the star(s) meeting those expectations, but in the end, only the star(s) received attention in the form of BSP from peers, so perhaps perceived as less effective by teachers. Since these studies had neutral or mixed evidence, it also could be that many students found being the ‘star’ and thus the center of attention at the end of the day or session was embarrassing or aversive—at least one student in PPR studies, “Katie” (Moroz and Jones, 2002), did better when she was the

praiser, not the recipient of peer praise (while some students would certainly desire to stand center-stage and have peer praises heaved upon them). These could be reasons why investigation shifted to tootling, where, in theory, the whole class would have better behavior as everyone can tootle on everyone. Possible downsides to tootling compared to PPR include the loss of students receiving that BSP directly from peers (because teachers read tootles aloud compared to PPR stars hearing BSP from peers) and how not all tootles are shared with the intended recipients when teachers only read 3–5 at the designated time. This might balance a limitation to PPR studies though, how students in PPR studies do not write down the good behavior they notice and might forget who and what they saw by the time it was PPR reporting. This lower dosage of BSP for the star in PPR studies might be comparable to how tootles are not all read to students.

Even with the shift to tootling, which allowed for classwide student recognition from peers, it was surprising that most students in tootling studies who were praised by peers on a tootle slip probably never knew

it. In almost all tootling studies, only 3–5 tootles were read at the end of the tootling period, class, or day, followed by all tootles being counted and the group reward tracker updated. What happened to the tootles after the few were read aloud was not reported in studies except for Thoele (2024), who sent tootles home *after* the group reward was met. Typical procedures therefore appear to be missing the important opportunity of letting all students hear or read the praise that was intended for them. It should have been a quick and easy step for teachers or a student leader to at least distribute tootles to the recipient if there was not time in the day to read them all. There were some exceptions to this in tootling studies, however, where all students were able to receive their peers' praise. For example, Harry et al. (2023) and Barahona (2010) did not read tootles aloud to the class but instead publicly posted all tootles on a bulletin board after class where students could read them the next day or gave them to students to keep, respectively. Teachers in Ray (2019) did read all tootles aloud, and it is possible the teacher in Cihak et al. (2009) read all tootles aloud during the 20 min allotted at the end of class, but it was not explicitly stated. In McHugh Dillon et al. (2019), students typed tootles into computer stations that immediately displayed them on an interactive whiteboard for everyone to read. In these cases, which happened to have the largest effect sizes for academic engagement, all students were not only able to read the praise intended for them specifically but could read all tootles given by any student. This might have provided additional reminders to students about what behaviors were expected and/or helped increase student motivation to meet expectations in the hope of receiving similar tootles themselves the next day.

We expected to find more peer praise note studies in theses and dissertations and were surprised to find the literature so clearly split between PPR and tootling. We thought more interventions would have taught students to say or write BSP statements immediately and directly to their peers, just like adult educators say or write specific praise for students. However, there were only three. Lund (2000) had fifth-grade peer monitors use BSP and give tokens to students contingent on quiet on-task behaviors on a fixed-interval schedule, with results showing engagement improved dramatically and disruptions decreased for both token earners and peer monitors. Two other studies utilized peer praise notes as previously reported in the 50-year map of BSP literature of Ennis et al. (2020): Kennedy et al. (2014) compared teacher-written and student-written praise notes during art class for grades 2–4 in a residential facility and found both worked equally well to reduce inappropriate behavior; Teerlink et al. (2017) implemented peer praise notes schoolwide at recess for an elementary school with 2–3 students per grade trained to be peer praisers, demonstrating peer praise notes were effective at reducing playground office discipline referrals. We hope future researchers will continue to investigate the effects of student-delivered specific praise notes, as there were not enough studies to evaluate the practice for EBP determination, but it appears to be a promising practice that has students directly and immediately recognizing appropriate and prosocial behavior of their peers using BSP.

We found it interesting to learn PPR was not at least a *potentially EBP* when it was PPR studies where students heard directly from peers what they did well to earn their specific praise, even though the reporting did occur at the end of the session or day (delayed reinforcement). We expected praise heard directly from peers to be more impactful compared to tootles read by teachers.

We acknowledge, of course, many PPR studies were missing the required minimum participants to meet Council for Exceptional Children (2014) QIs or used a design that did not allow for the possibility of three demonstrations of effect, so it still could be that PPR's direct sharing of BSP to peers is more impactful than having teachers read student praises. Given the limited number of PPR studies meeting QIs, it is difficult and perhaps not appropriate to compare PPR effect sizes to tootling study effect sizes—some are higher, some lower, some similar—there are just too few.

Something to consider regarding the EBP of tootling, is how do we know what the effective component(s) of the intervention are? We learn in each study the class tootling goal and how many times the class met the goal to know the dosage of tootles written for the class, but readers do not learn the dosage per student, not even target students (similarly in PPR studies, dosage of praise statements received by the star was not reported). It could be that popular students received the most tootles. It could be that dosage is not important because the key component might be *knowing* peers are watching for good behavior even if they do not fill out a tootle of even if your tootle is not read aloud. Future studies should report the dosage of tootles written *and received* (read aloud or posted for reading) for each target student and the average per day per student in the classroom. In addition to unknown dosage, we also cannot isolate if peer-specific praise is a key factor in tootling—it is the teacher who reads tootles, so it is possible students receiving praise perceive it as teacher attention even though it was written by a peer. Many of the early studies of PPR targeted socially isolated students or peers who needed to increase positive interactions with peers (e.g., Ervin et al., 1996; Moroz and Jones, 2002). It seems counter-intuitive, then, to have the intervention be so teacher-driven, limiting the potential for positive social interactions and praise directly between peers. Plus, the teacher typically praised the appropriate behavior mentioned in the tootle, and both students who wrote and received each tootle drawn, so it could be that teacher attention/praise is the most responsible for changes in student outcomes. Additionally, observed changes in student behavior may be partially attributed to the interdependent group contingency in each tootling study, and we cannot know to what degree. We do know group contingency interventions, especially in general education classrooms, are an evidence-based practice when What Works Clearinghouse standards were applied (Maggin et al., 2017). The group contingency component of tootling interventions, for some students, might be the strongest motivator for good behavior, more so than peer praise or teacher attention, and receiving the group reward may be the most reinforcing aspect for some students. We believe changes in student behavior during tootling studies are most likely a combination of teacher attention, peer attention, group contingency/reward, and knowing your peers are watching you to write a tootle that might be read later. Future researchers could run a component analysis study to more definitively determine active ingredients in tootling interventions, with and without group contingency, and/or compare typical tootling procedures to truly student-delivered BSP interventions where students immediately and directly praise peers when they observe targeted prosocial behavior.

We encourage readers to keep in mind we used two different methods to look at the effects of each study. For Council for Exceptional Children (2014) to determine whether a single-case research design study had *positive effects*, 75% of cases needed to

have a functional relation in the therapeutic direction, and if not, the study was classified as *neutral or mixed effects*. Of the two methods we used, this is the more conservative approach using visual analysis. In comparison, when we calculated BC-SMD, all participant data were used, which could result in an effect size that, if examined in isolation (e.g., without having CEC classification at hand), could seem to indicate overall positive results. For example, [Chenier \(2010\)](#) had two of three PPR students with positive outcomes and the third student with neutral results—when looking at all three students as a non-concurrent multiple baseline, there was not a functional relation. Yet, the BC-SMD estimated effect size was 0.52 (medium effect), likely due to the large level changes in the two students who had therapeutic outcomes plus the small increase in level for the one student with neutral results. A similar comparison can be made in another PPR study, [Moroz and Jones \(2002\)](#), as well as in tootling studies such as [Wright \(2019\)](#). [Wright \(2019\)](#) demonstrated a functional relation in two of three classrooms (66.7%) in their A-B-A-B design where 75% was needed for *positive results*, so the [Council for Exceptional Children \(2014\)](#) classification was *neutral or mixed evidence*; the BC-SMD estimated effect size was 0.46 for disruptive behavior (small effect) and 0.91 for academically engaged behavior (large effect) when all student data are considered in the examination of mean level changes despite the lack of a functional relation. We therefore suggest readers interpret BC-SMD effect sizes with caution and with overall CEC study designation in mind. This is in alignment with [Maggin et al. \(2017\)](#) recommendations, who also applied BC-SMD effect size estimates in their meta-analysis of single-case research design group contingency studies. The authors noted that a lot more investigation is required in terms of how researchers separate assessments of effect size and methodological rigor in single-case research, but that using parametric analysis and visual analysis together in systematic literature reviews and meta-analyses is supported.

Limitations

As with any literature review, it is possible, despite our best efforts to be systematic in our search, that we missed including some studies. We followed [Lane et al.'s \(2022\)](#) established guidelines for an exhaustive search to prevent missing articles and included theses and dissertations to best represent the full literature base on peer praise. Future researchers might additionally attempt to conduct forward ancestral searches of the included studies. Similarly, although all steps of our study review process after procurement of articles (i.e., QI coding, descriptive coding, study evidence-base practice classification, effect size calculations) were completed by two authors with high levels of reliability, there is always the possibility there was an error in our coding or that other researchers may interpret results differently. Thus, we encourage interpretations of our results regarding the student-delivered BSP body of literature be made with caution as readers keep these limitations in mind.

Educational implications

Teachers in schools where PBIS or Ci3T is implemented might consider implementing one of the versions of student-delivered BSP. The whole school might even try it as a Tier 1 prevention effort

that extends PBIS to the student level as peer praisers, where teachers get help from students implementing the low-intensity strategy of BSP as a positive reinforcement for meeting schoolwide behavior expectations. Or, teachers might notice many students in their classroom need support staying on task or engaging in more prosocial behavior and decide to implement a version of student-delivered BSP in their classroom only, such as tootling, all day or for a particular time of day where behavior is most challenging. If just one or two students are socially isolated and not being included by peers, in addition to reteaching appropriate social skills lessons for all students, teachers could implement PPR and make those students the 'star' at a higher rate than peers. Teaching students to specifically praise their peers with PPR or tootling would not take more than a few minutes each day, would not interfere with other teacher-delivered low-intensity strategies that support engagement and reduce disruptions, and may help teachers increase their classroom self-efficacy to keep all students in the room learning.

The delayed specific praise seen in PPR and tootling studies worked for almost all student participants but not everyone, so it might work for more and have even larger impacts if teachers taught students to praise peers directly and immediately (e.g., "Thanks for cleaning up the floor around all our group's desks, Robyn") for a targeted time of day when challenging behavior is known to occur most often, or even the full day. Teachers could then reinforce direct and immediate student praise with teacher-delivered BSP (e.g., "Jayson, I love how you thanked Robyn for cleaning up the whole group table") to encourage student BSP to occur more regularly. A recommended component often considered key to the powerful impact of BSP is immediacy ([Ennis et al., 2018](#)), so making the shift in the classroom to praising peers right away instead of waiting until the end of the hour or the day might help students stay even more on task with appropriate behavior. Similarly, when done authentically, praising the recipient directly might be more impactful compared to students telling the teacher what they saw (praise recipient hears it but not directly addressed to them) or writing down what they saw for the teacher to read to the class later (praise recipient learns about it from the teacher but does not hear it from the praiser). In most tootling studies, teachers only read 3–5 tootles, so most students did not hear if a peer recognized their prosocial behavior, whereas teaching students to praise peers directly and immediately would allow all students to hear the praise intended for them and thus be more reinforcing to the behavior being specifically praised.

Most included studies took place at elementary grade levels when young students seek teacher attention, so it might make more sense to study student-delivered BSP at the middle and high school levels. Adolescent students in secondary schools tend to seek peer attention more than adult attention, so perhaps peer praise is best suited for middle and high school settings where students already seek out peer approval. Future researchers should do more peer praise studies at the secondary level to test if adolescents are indeed more motivated by and reinforced by peer attention in the form of student-delivered BSP compared to elementary students who typically desire teacher attention.

Summary

We conducted an exhaustive systematic literature review on student-delivered BSP to peers and found 36 articles focused primarily on positive peer reporting (PPR) and tootling interventions. We used

Council for Exceptional Children (2014) standards for evidence-based practices to code included articles for quality indicators (QI) using a weighted 80% criterion and classified PPR in the *insufficient evidence* category and tootling in the *evidence-based practice* category. We calculated each eligible (80% QI met; QI 6.5 met; three or more cases in single-case research designs) study's effect size, either between-case standardized mean difference estimate (A-B-A-B withdrawal/reversal and multiple baseline designs), standardized average difference between successive observations (for one alternating treatment design), or Hedge's g (two group designs), then calculated a random-effects meta-analysis for PPR at 0.2254 (small effect), 1.0238 (large effect) for tootling, and 0.7408 for all eligible studies. Future researchers should (a) continue to investigate PPR with sufficient participants using methodologically sound research designs, (b) conduct tootling studies in middle and high school settings, (c) component analysis studies of the tootling intervention to determine active ingredients (e.g., teacher attention, peer praise, dosage of teacher and student praise), (d) conduct additional peer praise note studies to allow determination of evidence-base practice category, and (e) conduct studies where students across contexts are taught to directly and immediately recognize peer prosocial behavior using BSP.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

References

- References indicated with an asterisk (*) were included in the literature review.
- Achenbach, T. M. (1991). Manual for the Child Behavior Checklist and 1991 profile: Department of Psychiatry, University of Vermont.
- Allday, R. A., Hinkson-Lee, K., Hudson, T., Neilsen-Gatti, S., Kleinke, A., and Russel, C. S. (2012). Training general educators to increase behavior-specific praise: Effects on students with EBD. *Behav. Disord.* 37, 87–98. doi: 10.1177/019874291203700203
- * Alstead, K. J. (2022). *Investigating the effectiveness of the positive reinforcement components of tootling*. (Publication No. 29161176) [Doctoral dissertation, Minnesota State University, Mankato]. ProQuest Dissertations and Theses Global.
- * Barahona C. (2010). A comparison of public and private positive peer reporting in general education classrooms (Publication No. 29125567) [Doctoral dissertation, Louisiana State University and Agricultural & Mechanical College]. ProQuest Dissertations & Theses Global.
- Bracken, B. A. (1993). Assessment of Interpersonal Relations: Examiner's manual: PRO-ED.
- Briesch, A. M., and Chafouleas, S. M. (2009). Children's usage rating profile: University of Connecticut.
- Brophy, J. (1981). Teacher praise: A functional analysis. *Rev. Educ. Res.* 51, 5–32. doi: 10.3102/00346543051001005
- Caldarella, P., Larsen, R. A., Williams, L., and Wills, H. P. (2023). Effects of middle school teachers' praise-to-reprimand ratios on students' classroom behavior. *J. Posit. Behav. Interv.* 25, 28–40. doi: 10.1177/10983007211035185
- * Chaffee, R. K., Briesch, A. M., Volpe, R. J., Johnson, A. H., and Dudley, L. (2020). Effects of a class-wide positive peer reporting intervention on middle school student behavior. *Behav. Disord.* 45, 224–237. doi: 10.1177/0198742919881112
- Chafouleas, S. M., Briesch, A. M., Neugebauer, S. R., and Riley-Tillman, T. C. (2011). Usage Rating Profile–Intervention (Revised): University of Connecticut.
- * Chenier J. S. (2010). A treatment components analysis in positive peer reporting for socially withdrawn children (Publication No. 29122961) [Doctoral dissertation, Louisiana State University and Agricultural & Mechanical College]. ProQuest Dissertations & Theses Global.
- * Cihak D. F., Kirk E. R., Boon R. T. (2009). Effects of classwide positive peer "tootling" to reduce the disruptive classroom behaviors of elementary students with and without disabilities. *J. Behav. Educ.* 18, 267–278. doi: 10.1007/s10864-009-9091-8
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104
- Coie, J. D., Dodge, K. A., and Coppotelli, H. (1982). Dimensions and types of social status: A cross-age perspective. *Dev. Psychol.* 18, 557–570. doi: 10.1037/0012-1649.18.4.557
- Collins, T. A., Drevon, D. D., Brown, A. M., Villarreal, J. N., Newman, C. L., and Endres, B. (2020). Say something nice: A meta-analytic review of peer reporting interventions. *J. Sch. Psychol.* 83, 89–103. doi: 10.1016/j.jsp.2020.10.002
- Cooper, J. O., Heron, T. E., and Heward, W. L. (2020). Applied behavior analysis. 3rd Edn: Pearson.
- Council for Exceptional Children (2014). Council for exceptional children standards for evidence-based practices in special education: Council for Exceptional Children.
- Dettoni, J. R., Norvell, D. C., and Chapman, J. R. (2022). Fixed-effect vs random-effects models for meta-analysis: 3 points to consider. *Global Spine J* 12, 1624–1626. doi: 10.1177/21925682221110527
- Ebesutani, C., Regan, J., Smith, A., Reise, S., Higa-McMillan, C., and Chorpita, B. F. (2012). The 10-item positive and negative affect schedule for children, child and parent shortened versions: Application of item response theory for more efficient assessment. *J. Psychopathol Behav Assess* 34, 191–203. doi: 10.1007/s10862-011-9273-2
- Ennis, R. P., and Losinski, M. (2019). Interventions to improve fraction skills for students with disabilities: A meta-analysis. *Except. Child.* 85, 367–386. doi: 10.1177/0014402918817504
- Ennis, R. P., Royer, D. J., Lane, K. L., and Dunlap, K. D. (2020). Behavior-specific praise in pk–12 settings: Mapping the 50-year knowledge base. *Behav. Disord.* 45, 131–147. doi: 10.1177/0198742919843075
- Ennis, R. P., Royer, D. J., Lane, K. L., and Griffith, C. (2017). A systematic review of precorrection in pk–12 settings. *Educ. Treat. Child.* 40, 465–495. doi: 10.1353/etc.2017.0021
- Ennis, R. P., Royer, D. J., Lane, K. L., Menzies, H. M., Oakes, W. P., and Schellman, L. E. (2018). Behavior-specific praise: An effective, efficient, low-intensity strategy to support student success. *Beyond Behav* 27, 134–139. doi: 10.1177/1074295618798587
- * Ervin R. A., Miller P. M., Friman P. C. (1996). Feed the hungry bee: Using positive peer reports to improve the social interactions and acceptance of a socially rejected girl in residential care. *J. Appl. Behav. Anal.* 29, 251–253. doi: 10.1901/jaba.1996.29-251

Author contributions

DR: Writing – original draft, Writing – review & editing. RE: Writing – original draft, Writing – review & editing.

Funding

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

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Forness, S. R., Freeman, S. F., Paparella, T., Kauffman, J. M., and Walker, H. M. (2012). Special education implications of point and cumulative prevalence for children with emotional or behavioral disorders. *J. Emot. Behav. Disord.* 20, 4–18. doi: 10.1177/1063426611401624
- Fritz, C. O., Morris, P. E., and Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *J. Exp. Psychol. Gen.* 141, 2–18. doi: 10.1037/a0024338
- Furlong, M. J., You, S., Renshaw, T. L., O'Malley, M. D., and Rebelez, J. (2013). Preliminary development of the Positive Experiences at School Scale for elementary school children. *Child Indic. Res.* 6, 753–775. doi: 10.1007/s12187-013-9193-7
- Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *J. Am. Acad. Child Adolesc. Psychiatry* 40, 1337–1345. doi: 10.1097/00004583-200111000-00015
- * Gray J. L. (2023). The integration of positive psychology and positive behavioral interventions & supports to improve minoritized students' social, emotional, and behavioral outcomes (Publication No. 29258155) [Doctoral dissertation, University of South Florida]. ProQuest Dissertations & Theses Global.
- Gresham, F. M., Cook, C. R., Collins, T., Dart, E., Rasetshwane, K., Truelson, E., et al. (2010). Developing a change-sensitive brief behavior rating scale as a progress monitoring tool for social behavior: An example using the social skills rating system—teacher form. *School Psychol Rev* 39, 364–379. doi: 10.1080/02796015.2010.12087758
- Gresham, F., and Elliott, S. N. (2008). Social skills improvement system (SSIS) rating scales: Pearson Assessments.
- * Grieger T., Kauffman J. M., Grieger R. M. (1976). Effects of peer reporting on cooperative play and aggression of kindergarten children. *J. Sch. Psychol.* 14, 307–313. doi: 10.1016/0022-4405(76)90027-3
- * Harry S. W., Tingstrom D. H., Dufrene B. A., Dart E. H., Radley K. C., Lum J. D., et al. (2023). The effects of tootling combined with public posting in high school classrooms. *J. Behav. Educ.* 32, 1–20. doi: 10.1007/s10864-021-09462-1
- * Hoff K., Ronk M. (2006). Increasing prosocial interactions using peers: Extension of positive peer-reporting methods. *J. Evid Based Pract Schools* 7, 27–42.
- Huebner, E. S. (1991). Initial development of the Student's Life Satisfaction Scale. *School Psychol Int* 12, 231–240. doi: 10.1177/0143034391123010
- Huebner, E. S., and Gilman, R. (2002). An introduction to the multidimensional students' life satisfaction scale. *Soc Indic Res* 60, 115–122. doi: 10.1023/A:1021252812882
- Individuals with Disabilities Education Improvement Act. (2004). Chapter 20 of the United States Code section 1400 et seq.
- Irwin, D. E., Stucky, B., Langer, M. M., Thissen, D., DeWitt, E. M., Lai, J. S., et al. (2010). An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. *Qual. Life Res.* 19, 595–607. doi: 10.1007/s11136-010-9619-3
- * Johnson, C. (2008). *Effects of a positive peer reporting intervention on prosocial interactions in a general education classroom.* (Publication No. 3371515) [Doctoral dissertation, Georgia State University]. ProQuest Dissertations & Theses Global.
- * Jones, K. M., Young, M. M., and Friman, P. C. (2000). Increasing peer praise of socially rejected delinquent youth: Effects on cooperation and acceptance. *Sch. Psychol. Q.* 15, 30–39. doi: 10.1037/s0088776
- Kelley, M. L., Heffer, R. W., Gresham, F. M., and Elliott, S. N. (1989). Development of a modified treatment evaluation inventory. *J. Psychopathol Behav Assess* 11, 235–247. doi: 10.1007/BF00960495
- * Kennedy C., Jolivet K., Ramsey M. L. (2014). The effects of written teacher and peer praise notes on the inappropriate behaviors of elementary students with emotional and behavioral disorders in a residential school. *Resid. Treat. Child. Youth* 31, 17–40. doi: 10.1080/0886571X.2014.878577
- * Kirkpatrick B. A., Wright S., Daniels S., Taylor K. L., McCurdy M., Skinner C. H. (2019). Tootling in an after-school setting: Decreasing antisocial interactions in at-risk students. *J. Posit. Behav. Interv.* 21, 228–237. doi: 10.1177/1098300719851226
- Ladd, G. W., and Proffitt, S. M. (1996). The Child Behavior Scale: A teacher-report measure of children's aggressive, withdrawn, and prosocial behaviors. *Dev. Psychol.* 32, 1008–1024. doi: 10.1037/0012-1649.32.6.1008
- Lajeunesse, M. J. (2021). *Creating publication-quality forest plots in microsoft excel.* [Online resource]. doi: 10.6084/m9.figshare.14161310
- * Lambert A. M. (2014). Evaluating the use of tootling for improving upper elementary/middle school students' disruptive and appropriate behavior (Publication No. 3583919) [Doctoral dissertation, The University of Southern Mississippi]. ProQuest Dissertations & Theses Global.
- * Lambert A. M., Tingstrom D. H., Sterling H. E., Dufrene B. A., Lynne S. (2015). Effects of tootling on classwide disruptive and appropriate behavior of upper-elementary students. *Behav. Modif.* 39, 413–430. doi: 10.1177/0145445514566506
- Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174. doi: 10.2307/2529310
- Lane, K. L., Common, E. A., Royer, D. J., and Muller, K. (2019a). Group comparison and single-case research design quality indicator matrix using council for exceptional children 2014 standards. Available online at: <https://www.ci3t.org/practice>
- Lane, K. L., Common, E. A., Royer, D. J., and Oakes, W. P. (2022). "Conducting systematic reviews of the literature: Guidance for quality appraisals" in *Advances in learning and behavioral disabilities.* eds. M. J. Tankersley, B. G. Cook and T. J. Landrum, vol. 32 (Emerald Insight), 109–130.
- Lane, K. L., Kalberg, J. R., and Shepcaro, J. C. (2009). An examination of the evidence base for function-based interventions for students with EBD attending middle and high schools. *Except. Child.* 75, 321–340. doi: 10.1177/001440290907500304
- Lane, K. L., Menzies, H. M., Ennis, R. P., and Oakes, W. P. (2015). Supporting behavior for school success: A step-by-step guide to key strategies: Guilford Press.
- Lane, K. L., Oakes, W. P., Cantwell, E. D., and Royer, D. J. (2019b). Building and installing comprehensive, integrated, three-tiered (Ci3T) models of prevention: A practical guide to supporting school success (v1.3): KOI Education.
- Lanza, G. E. (2020). Teachers' self-efficacy when managing disruptive student behaviors and its influence on teacher burnout (Publication No. 2406) [Doctoral dissertation, University of Connecticut].
- * Libster, L. (2009). *The efficacy of positive peer reporting procedures for use with neglected-status students in general education classrooms.* (Publication No. 29125864) [Doctoral dissertation, Louisiana State University and Agricultural & Mechanical College]. ProQuest Dissertations & Theses Global.
- * Lum J. D., Radley K. C., Tingstrom D. H., Dufrene B. A., Olmi D. J., Wright S. J. (2019). Tootling with a randomized independent group contingency to improve high school classwide behavior. *J. Posit. Behav. Interv.* 21, 93–105. doi: 10.1177/1098300718792663
- * Lum J. D., Tingstrom D. H., Dufrene B. A., Radley K. C., Lynne S. (2017). Effects of tootling on classwide disruptive and academically engaged behavior of general-education high school students. *Psychol. Sch.* 54, 370–384. doi: 10.1002/pits.22002
- * Lund C. J. (2000). The effects of a peer-mediated intervention on the classroom behaviors of token earners and peer helpers: An investigation of stimulus generalization (Publication No. 9980750) [Doctoral dissertation, The Florida State University]. ProQuest Dissertations & Theses Global.
- Maggin, D. M., Pustejovsky, J. E., and Johnson, A. H. (2017). A meta-analysis of school-based group contingency interventions for students with challenging behavior: An update. *Remedial Spec. Educ.* 38, 353–370. doi: 10.1177/0741932517716900
- Manolov, R., and Onghena, P. (2018). Analyzing data from single-case alternating treatments designs. *Psychol. Methods* 23, 480–504. doi: 10.1037/met0001133
- Martens, B. K., Witt, J. C., Elliott, S. N., and Darveaux, D. X. (1985). Teacher judgments concerning the acceptability of school-based interventions. *Profess Psychol* 16, 191–198. doi: 10.1037/0735-7028.16.2.191
- McDermott, P. A., Marston, N. C., and Stott, D. H. (1993). National profile in youth psychopathology: Adjustment scales for children and adolescents: Edumetric and Clinical Science.
- * McHugh Dillon M. B., Radley K. C., Tingstrom D. H., Dart E. H., Barry C. T. (2019). The effects of tootling via classdojo on student behavior in elementary classrooms. *Sch. Psychol. Rev.* 48, 18–30. doi: 10.17105/SPR-2017-0090.V48-1
- * McHugh M. B., Tingstrom D. H., Radley K. C., Barry C. T., Walker K. M. (2016). Effects of tootling on classwide and individual disruptive and academically engaged behavior of lower-elementary students. *Behav. Interv.* 31, 332–354. doi: 10.1002/bin.1447
- Menzies, H. M., Chang, Y.-C., and Smith-Menzies, L. (2023). Using behavior-specific praise to support intrinsic motivation. *Beyond Behav* 32, 173–181. doi: 10.1177/10742956231200899
- * Moroz K. B., Jones K. M. (2002). The effects of positive peer reporting on children's social involvement. *Sch. Psychol. Rev.* 31, 235–245. doi: 10.1080/02796015.2002.12086153
- * Morrison J. Q., Jones K. M. (2007). The effects of positive peer reporting as a class-wide positive behavior support. *J. Behav. Educ.* 16, 111–124. doi: 10.1007/s10864-006-9005-y
- Mundschen, D., and Simpson, R. (2014). "Defining emotional or behavioral disorders: The quest for affirmation" in *The sage handbook of emotional and behavioral difficulties.* eds. P. Garner, J. M. Kauffman and J. Elliott. 2nd ed (SAGE), 43–54.
- * Murphy J. K. (2013). Positive peer reporting interventions with high vs. Low experimenter-involvement in an alternative education setting (Publication No. 1536905) [Doctoral dissertation, University of South Alabama]. ProQuest Dissertations & Theses Global.
- O'Connor, E. E., Dearing, E., and Collins, B. A. (2011). Teacher-child relationship and behavior problem trajectories in elementary school. *Am. Educ. Res. J.* 48, 120–162. doi: 10.3102/0002831210365008
- Oden, S. (1980). "A child's social isolation: Origins, prevention, intervention" in *Teaching social skills to children.* eds. G. Cartledge and J. F. Milburn (Pergamon Press), 179–202.
- Peterson Nelson, J. A., Caldarella, P., Young, K. R., and Webb, N. (2008). Using peer praise notes to increase the social involvement of withdrawn adolescents. *Teach. Except. Child.* 41, 6–13. doi: 10.1177/004005990804100201
- Pianta, R. C. (2001). Student-Teacher Relationship Scale-Short Form. Psychological Assessment Resources, Inc. Available online at: <http://curry.virginia.edu/about/directory/robert-c.-pianta/measures>

- Pustejovsky, J. E., Chen, M., Hamilton, B., and Grekov, P. (2023). Scdhlm: A web-based calculator for between-case standardized mean differences (version 0.7.2) [web application]. Available online at: <https://jepusto.shinyapps.io/scdhlm>
- * Ray J. E. (2019). Evaluating the effects of tootling when implemented in special education classrooms providing behavior supports (Publication No. 27548214) [Doctoral dissertation, Minnesota State University, Mankato]. ProQuest Dissertations & Theses Global.
- Renshaw, T. L., Long, A. C. J., and Cook, C. R. (2015). Assessing adolescents' positive psychological functioning at school: Development and validation of the student subjective wellbeing questionnaire. *School Psychol Q* 30, 534–552. doi: 10.1037/spq0000088
- Rohatgi, A. (2024). Webplotdigitizer (version 4.8) [web-based application]. Available online at: <https://automeris.io/WebPlotDigitizer>
- Romer, N., von der Embse, N. P., Eklund, K., Kilgus, S. P., Perales, K., Splett, J. W., et al. (2020). Best practices in social, emotional, and behavioral screening: An implementation guide. Version 2.0. School Mental Health Collaborative. Available online at: <https://smhcollaborative.org/universalscreening/>
- Royer, D. J., Lane, K. L., Cantwell, E. D., and Messenger, M. L. (2017). A systematic review of the evidence base for instructional choice in k-12 settings. *Behav. Disord.* 42, 89–107. doi: 10.1177/0198742916688655
- Royer, D. J., Lane, K. L., Dunlap, K. D., and Ennis, R. P. (2019). A systematic review of teacher-delivered behavior-specific praise on k-12 student performance. *Remedial Spec. Educ.* 40, 112–128. doi: 10.1177/0741932517751054
- * Salinas, E. A. (2021). *Evaluating the effects of tootling on disruptive and academically engaged behaviors of students with emotional and behavioral disorders*. (Publication No. 28496210) [Doctoral dissertation, The University of Texas at San Antonio]. ProQuest Dissertations & Theses Global.
- Schluter, D. (2024). Meta-analysis. University of British Columbia. Available online at: <https://www.zoology.ubc.ca/~schluter/R/Meta.html> (Accessed June 1, 2024).
- Scott, T. M., Hirn, R. G., and Cooper, J. T. (2017). Teacher and student behaviors: Keys to success in classroom instruction: Rowman & Littlefield.
- * Shelton H. L. (2002). Increasing students' awareness and perception of peer prosocial behavior: An investigation of tootling (Publication No. 3059769) [Doctoral dissertation, The University of Tennessee]. ProQuest Dissertations & Theses Global.
- * Sherman J. C. (2012). Positive peer reporting and positive peer reporting combined with tootling: A comparison of interventions (Publication No. 3530739) [Doctoral dissertation, The University of Southern Mississippi]. ProQuest Dissertations & Theses Global.
- * Short M. B. (1999). Evaluating the effects of positive peer reporting on social acceptance (Publication No. 9951243) [Doctoral dissertation, Western Michigan University]. ProQuest Dissertations & Theses Global.
- Skinner, C. H., Skinner, A. L., and Cashwell, T. H. (1998). Tootling, not tattling [Paper presentation]. In: *26th Annual Mid South Educational Research Association, New Orleans, LA*.
- * Smith, S. M., Simon, J., and Bramlett, R. K. (2009). Effects of positive peer reporting (PPR) on social acceptance and negative behaviors among peer-rejected preschool children. *J. Appl. Sch. Psychol.* 25, 323–341. doi: 10.1080/15377900802595359
- * Steeves R. M. O. (2017). Differential effects of gratitude and positive peer reporting interventions on student subjective wellbeing, classroom behavior, and school connectedness (Publication No. 29117957) [Doctoral dissertation, Louisiana State University and Agricultural & Mechanical College]. ProQuest Dissertations & Theses Global.
- Sugai, G., and Horner, R. H. (2019). Sustaining and scaling positive behavioral interventions and supports: Implementation drivers, outcomes, and considerations. *Except. Child.* 86, 120–136. doi: 10.1177/0014402919855331
- Suldo, S. M. (2016). Promoting student happiness: Positive psychology interventions in schools: Guilford Press.
- Sutherland, K. S., Wehby, J. H., and Copeland, S. R. (2000). Effect of varying rates of behavior-specific praise on the on-task behavior of students with EBD. *J. Emot. Behav. Disord.* 8, 2–8. doi: 10.1177/10634266000800101
- Tanol, G., Johnson, L., McComas, J., and Cote, E. (2010). Responding to rule violations or rule following: A comparison of two versions of the good behavior game with kindergarten students. *J. Sch. Psychol.* 48, 337–355. doi: 10.1016/j.jsp.2010.06.001
- * Teerlink E., Caldarella P., Anderson D. H., Richardson M. J., Guzman E. G. (2017). Addressing problem behavior at recess using peer praise notes. *J. Posit. Behav. Interv.* 19, 115–126. doi: 10.1177/1098300716675733
- * Thoele J. M. (2024). The effects of student-and teacher-led tootling on students with disruptive behavior [Doctoral dissertation, University of Georgia].
- U.S. Department of Education. (2022). 44th annual report to congress on the implementation of the individuals with disabilities education act. Office of Special Education and Rehabilitative Services. Available online at: <http://www.ed.gov/about/reports/annual/osep>
- Valentine, J. C., Tanner-Smith, E. E., Pustejovsky, J. E., and Lau, T. S. (2016). Between-case standardized mean difference effect sizes for single-case designs: a primer and tutorial using the scdhlm web application. *Campbell Syst. Rev.* 12, 1–31. doi: 10.4073/cmdp.2016.1
- Von Brock, M. B., and Elliott, S. N. (1987). Influence of treatment effectiveness information on the acceptability of classroom interventions. *J. Sch. Psychol.* 25, 131–144. doi: 10.1016/0022-4405(87)90022-7
- Walker, H. M., and Severson, H. (1992). Systematic screening for behavior disorders: User's guide and technical manual: Sopris West.
- Wilson, P., Rhymer, K., Landis, J., and Skinner, C. (2001). Promoting self-concept, social skills, and interpersonal relations: The tootling intervention [Paper presentation]. In: *Annual Conference of the National Association of School Psychologists, Washington, DC*.
- Witt, J. C., and Elliott, S. N. (1985). "Acceptability of classroom intervention strategies" in *Advances in school psychology*. ed. T. R. Kratochwill (Erlbaum), 4, 251–288.
- * Wright S. J. (2019). Peer-mediated tootling with a standardized form and a mystery motivator in high school classrooms (Publication No. 10751872) [Doctoral dissertation, The University of Southern Mississippi]. ProQuest Dissertations & Theses Global.
- Zweers, I., van de Schoot, R. A., Tick, N. T., Depaoli, S., Clifton, J. P., de Castro, B. O., et al. (2021). Social-emotional development of students with social-emotional and behavioral difficulties in inclusive regular and exclusive special education. *Int. J. Behav. Dev.* 45, 59–68. doi: 10.1177/0165025420915527