Investigating strategies to increase general education teachers' adherence to evidence-based social-emotional behavior practices: A meta-analysis of the single-case literature.

James L. Merle[1], Andrew J. Thayer[2], Madeline F. Larson[3], Sydney Pauling[4], Clayton R. Cook[3], Joseph A. Rios[3], Jenna L. McGinnis[3], Margaret M. Sullivan[3]

[1] University of Utah

[2] Thayer Organizational & Psychological Services, LLC

[3] University of Minnesota - Twin Cities

[4] University of Texas at Austin

First published online December 24, 2021

Merle, J. L., Thayer, A. J., Larson, M. F., Pauling, S., Cook, C. R., Rios, J. A., McGinnis, J. L., Sullivan, M. (2021). Investigating strategies to increase general education teachers' adherence to evidence-based social-emotional behavior practices: A meta-analysis of the single-case literature. Journal of School Psychology, 91, 1-26. https://doi.org/10.1016/j.jsp.2021.11.005

merle016@umn.edu; james.merle@.utah.edu; Address: Department of Population Health Sciences, University of Utah School of Medicine, 295 Chipeta Way, Salt Lake City, UT 84108

Abstract

Educational researchers have produced a variety of evidence-based practices (EBP) to address social, emotional, and behavioral (SEB) needs among students. Yet, these practices are often insufficiently adopted and implemented with fidelity by teachers to produce the beneficial outcomes associated with the EBP, leaving students at risk for developing SEB problems. If ignored, SEB problems can lead to other negative outcomes, such as academic failure. Therefore, implementation strategies (i.e., methods and procedures designed to promote implementation outcomes) are needed to improve teachers' uptake and delivery of EBPs with fidelity. This meta-analysis sought to examine the types and magnitude of effect of implementation strategies that have been designed and tested to improve teacher adherence to SEB EBPs. Included studies (a) used single case experimental designs, (b) employed at least one implementation strategy, (c) targeted general education teachers, and (d) evaluated adherence as a core dimension of fidelity related to the delivery of EBPs. In total, this study included 28 articles and evaluated 122 effect sizes. A total of 15 unique implementation strategies were categorized. Results indicated that, on average, implementation strategies were associated with increases in teacher adherence to EBPs above baseline and group-based pre-implementation trainings alone (g = 2.32, tau = .77). Moderator analysis also indicated that larger effects were associated with implementation strategies that used a greater number of unique behavior change techniques ($p < .001$). Implications and future directions for research and practice regarding use of implementation strategies for general education teachers are discussed.

Keywords. Implementation strategies, school-based, single case, meta-analysis, mechanism, behavior change

Investigating strategies to increase general education teachers' adherence to evidence-based social-emotional behavior practices: A meta-analysis of the single-case literature

A wide range of evidence-based practices (EBP) have been developed to address student social, emotional, and behavioral (SEB) needs (Owens et al., 2014). General education teachers are the primary implementers of universal SEB EBPs designed for all students in a school or classroom, which aim to prevent SEB problems and promote success-enabling factors (Reinke et al., 2011). Teachers also play a key role in implementing specific practices associated with targeted and intensive interventions for students who have needs that surpass universal supports. Teacher delivery of the core components of an EBP with fidelity (i.e., implementing as it was designed) is a critical determinant of promoting student outcomes (Biggs et al., 2008). Given the importance of fidelity, researchers have developed and tested implementation strategies (i.e., methods and procedures aimed at improving specific implementation outcomes) to improve general education teachers' adherence to EBPs (e.g., Collier-Meek et al., 2016; Dart, et al., 2012; Sanetti & Kratochwill, 2011). Although existing systematic reviews have examined the efficacy of implementation strategies, no meta-analysis to date has categorized and examined implementation strategies in accordance with established reporting standards to capture key characteristics that may be associated with efficacy (e.g., actor, temporality; Proctor et al., 2011) nor attempted to distill strategies down into the specific behavior change techniques (BCTs) used and the underlying mechanisms of action being targeted (MoA; Michie et al., 2013). Thus, in this meta-analysis, we sought to (a) synthesize the extant implementation strategies used across the single case literature to support general education teachers' delivery of EBPs with fidelity, (b) explore the magnitude of effects associated with particular implementation strategies, and (c) determine whether certain variables (i.e., the number of BCTs used, the intervention tier which

the SEB EBP was delivered, the grade each teacher taught, study quality indicators, and the total years of experience) moderate the impact of strategies on teacher adherence.

Social, Emotional, and Behavior Needs and Evidence-Based Practices

Schools are expected to promote the development of SEB skills and competencies (Domitrovich et al., 2017; Greenberg et al., 2003). Students who experience SEB problems impede teaching and learning (Cook et al., 2013), negatively impact social relationships (Cook et al., 2010), are likely to receive exclusionary discipline (Skiba & Rausch, 2006), and have higher rates of academic failure and dropout (Snyder & Dillow, 2015). Youth who drop out are at increased risk for negative outcomes later in life, such as unemployment and lower economic earnings in adulthood (Simmons, 2013), reliance on public assistance (Waldfogel et al., 2007), substance abuse (Townsend et al., 2007), and incarceration (Moore et al., 2013).

A variety of programs exist to promote SEB skills. Some programs promote students' ability to display prosocial behaviors and make positive choices through enhancing self-esteem (Catalano et al., 2003), whereas others teach a range of social-emotional learning (SEL) skills and competencies, such as the ability to recognize and manage emotions, solve problems, and build positive relationships with others (Blewitt et al., 2018; Collaborative for Academic, Social, and Emotional Learning, 2012; Taylor et al., 2017). Furthermore, developing SEB skills is associated with academic gains (Durlak et al., 2011). Thus, stakeholders have pushed for SEL programs to be integrated into school practices alongside academics to promote overall student success (Brackett & Rivers, 2014). Consistent with this goal, a host of EBPs to address SEB needs have been developed for implementation in schools.

EBP is an umbrella term that has been defined as "the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and

preferences" (American Psychological Association [APA] Presidential Task Force on Evidence-Based Practice, 2006, p. 273). EBPs include not only treatments or interventions, but also assessments, structures, and strategies (e.g., data-based decision making; Kowalski & Lasley, 2009). Determining EBPs in education includes a systematic approach to determining which programs, interventions, or procedures are supported by a sufficient number of studies that have high methodological quality, use appropriate research designs that allow for an assessment of efficacy, and demonstrate meaningful change amongst a generalizable sample of participants (Cook et al., 2012). To identify SEB EBPs in this study, we used the standards provided by What Works Clearinghouse (WWC), an arm of the Institute for Education Sciences within the U.S. Department of Education, which has established "gold standard" requirements for evidence-based practices (i.e., study methodological quality; WWC, 2020).

EBP Delivery in Schools

In recognition of the impact SEB competencies have on short- and long-term outcomes, professional organizations (e.g., APA, National Association of School Psychologists) and federal legislation (e.g., Individuals with Disabilities Education Improvement Act, 2004; Every Students Succeeds Act, 2015) call for the adoption and delivery of EBPs that aim to prevent and address SEB needs. Most schools in the U.S. organize and deliver EBPs through a multi-tiered system of support (MTSS), which is often conceptualized as a three-tier framework that facilitates the delivery of a continuum of supports matched to the level of student need (Buckle, n.d.; Cook et al., 2010). Unfortunately, researchers have consistently shown that an implementation gap exists, with numerous barriers impeding the successful translation of EBPs into routine practice in schools, leading to uneven, inconsistent, and incomplete implementation (Durlak & DuPre, 2008; Krause et al., 2014).

To address this gap, researchers have increasingly shifted their attention from developing and testing EBPs to examining factors and strategies associated with effectively transporting existing EBPs into practice. Some researchers contend that an EBP saturation point has been reached (Chorpita & Daleiden, 2009; Kessler & Glasgow, 2011; Lyon & Bruns, 2019a) and that an "implementation perspective" (Petersilia, 1990, p. 129) is needed that shifts research priorities away from the development of new innovations to focusing attention on the processes involved to support the successful implementation of existing EBPs.

General Education Teachers and EBP Implementation. General education is a setting designed for all children. It is where students spend most of their time during school hours and where core subjects are taught and evaluated (Webster, 2019). General education teachers play a critical role in the implementation of EBPs within a multi-tiered system of support. For many universal SEB EBPs, general education teachers are the primary implementers because of their proximity to and regular face-to-face interactions with students (Reinke et al., 2011).

Indeed, there are factors at various socio-ecological levels that impact teachers' decisions to implement a given practice, such as outer and inner setting factors including leadership, climate, and funding sources (Lyon & Bruns, 2019b). Teachers need to have a supportive infrastructure that facilitates their implementation of EBPs and reduces common barriers; these barriers include lack of protected time or capacity, unsupportive administrators, lack of buy-in, and turnover (Turri et al., 2016). Implementation strategies exist to target outer and inner setting factors, such as leadership training (Aarons et al., 2015), planning for and building capacity, planning for annual onboarding of new staff, involving teachers as key stakeholders in implementation decisions, and clarifying how proposed changes align with other initiatives to foster staff buy-in (George et al., 2009). However, it is ultimately up to front line implementers,

such as teachers, to overcome barriers and decide which SEB practices they implement (Locke et al., 2019; Ramos-Vidal et al., 2020).

General education teachers have an observable influence on student SEB outcomes through the practices they adopt and deliver (George et al., 2009). Effective practices include establishing clear expectations for behavior, teaching social-emotional competencies through a curriculum-led process, developing systems to recognize and acknowledge students for exhibiting expected behavior, effectively responding to problem behavior with compassion and empathy, and fostering positive classroom climate through cultivating positive relationships (Cook et al., 2018). At subsequent tiers of support, general education teachers are often responsible for delivering specific components of an intervention (Noell, Gresham, & Gansle, 2002). As co-implementers of interventions, general education teachers deliver practices to support the generalization and maintenance of skills learned in the context of small group interventions (i.e., precorrection, prompts, and positive reinforcement) or deliver proactive or reactive strategies included as part of a multi-pronged behavior intervention plan (Hawken et al., 2009).

For students to benefit from the variety of available EBPs, it is crucial that general education teachers deliver EBPs as designed and as shown to be effective (i.e., with fidelity). It is also important to consider adaptations to existing interventions needed to match the population served by the school as well as with available resources (Fernandez et al., 2019). Large group trainings and didactics are cornerstone implementation strategies used widely by schools to launch new initiatives (Lyon et al., 2017). However, research has shown that although these trainings can be effective for some implementers, only around 40% ultimately implement new practices without follow-up supports (O'Donnell, 2008; Sanetti et al., 2013), and among teachers

who successfully adopt, adherence to core treatment practices has been shown to fade within 10 days of beginning implementation without ongoing support (Sanetti & Kratochwill, 2009b; Reinke et al., 2008). When explaining this problem, prior implementation research has identified a variety of barriers to implementation, including lack of motivation, self-efficacy, protected time, and administrative support (Collier-Meek et al., 2017; Long et al., 2016; Ransford et al., 2009). Furthermore, low fidelity to EBPs reduces the likelihood that students can benefit from them and experience improved SEB outcomes (Reinke et al., 2011). Researchers from the field of teacher professional development also indicate a need for more research on efficient and effective implementation supports for teachers around practices promoting SEB needs (Darling-Hammond et al., 2017; State et al., 2019). Therefore, implementation research is needed that specifies *what* works to support teachers' adoption and delivery of EBPs and *how* and *why* those effects are achieved.

Implementation Science

The limited transfer of EBPs into routine practice is experienced across a range of child-serving settings (e.g., healthcare, child welfare, education) and has led to the rapid emergence of the multidisciplinary field of implementation science. Implementation researchers have identified numerous plausible determinants (i.e., factors that enable or obstruct successful implementation) across multiple levels of influence (Krauss et al., 2014), including (a) outer setting determinants (i.e., factors beyond the immediate setting in which implementation takes place, such as policy, district leadership, and funding), (b) inner setting determinants (i.e., the specific setting in which implementation takes place, such as principal leadership qualities, teacher champions, and school climate), (c) determinants associated with individuals who are expected to adopt and deliver the practice (e.g., teachers' beliefs, attitudes, intentions, burnout), and (d) determinants associated

with the practice or intervention itself (i.e., how feasible, appropriate, and acceptable a practice is; Lyon & Bruns, 2019a).

Despite the importance of outer and inner setting factors, implementation of EBPs rests with the decisions and actions of the implementers who receive training and are expected to adopt, deliver, and sustain a given EBP. For example, Locke and colleagues (2019) compared the impact of individual teacher attitudes and organizational factors, such as leadership and climate, and found that individual attitudes impacted implementation more so than organization factors. Indeed, individual-level characteristics of front-line implementers have been consistently argued as central to implementation success (e.g., Tabak et al., 2012). Two key areas of research within implementation science include studying the implementation *outcomes* that promote the beneficial effects of services (i.e., whether the innovation was delivered as intended by its developers), which are attained through the development and testing of implementation *strategies* (Powell et al., 2012).

Implementation Outcomes

Implementation outcomes are defined as "…the effects of deliberate and purposive actions to implement new treatments, practices, and services" (Proctor et al., 2011, p. 65). Implementation outcomes are adult-facing, concerned with whether effective practices are adopted as planned—compared to effectiveness or efficacy outcomes, which are child-facing, and are concerned with whether the practice led to improvements in functioning of the students, patients, or clients receiving the practice. There are eight core implementation outcomes as defined by Proctor et al.'s (2011) seminal study. These include (a) acceptability, (b) adoption, (c) appropriateness, (c) costs, (d) feasibility, (e) fidelity, (f) penetration, and (g) sustainability.

In terms of individual-level implementation outcomes, fidelity is one of the most

important due to research showing the relationship between fidelity and distal student outcomes (O'Donnell, 2008). Fidelity is defined as the degree to which core components of a practice are delivered as intended (Sanetti & Kratochwill, 2007). Fidelity has five dimensions: (a) *adherence* to the program protocol (i.e., whether the core components or activities of program or intervention is being delivered as it was designed), (b) *dosage/exposure* reflecting the amount of the program or practice that was delivered and received, (c) *quality* of program delivery (i.e., skill in using the techniques or methods prescribed by the program, enthusiasm, preparedness, and attitude), (d) *program differentiation* (i.e., the unique features of different components or programs that are reliably differentiated from one another), and (e) *participant responsiveness* (i.e., the extent to which participants are engaged by and involved in receiving the program; Proctor et al., 2011). Among these, adherence is the most heavily researched dimension, as it is often considered a pre-condition for the others (McKenna et al., 2014; Schulte et al., 2009). In this study, we focused on adherence solely because most studies used adherence as their main and only outcome.

A positive relationship exists between adherence to the core components of EBPs and child outcomes across diverse SEB interventions (Blewitt et al., 2018; Durlak & DuPre, 2008). Logically speaking, students cannot benefit from support they do not receive. It is well documented that many general education teachers struggle to adhere to delivering core components of EBPs as intended (Sanetti et al., 2014) and evidence suggests that when provided with similar implementation supports, general education teachers implement EBPs at significantly lower rates than special education teachers (Solomon et al., 2012). There are several barriers that interfere with teacher adherence and thus limit the potential of EBPs to promote student outcomes. Teachers may have limited protected time, lack knowledge or motivation,

experience stress and exhaustion due to initiative overload, and receive insufficient follow-up support to deliver EBPs with integrity (Larson et al., 2018; Myers et al., 2011; Putnam et al., 2009). For these reasons, implementation strategies are needed to promote general education teachers' adherence to EBPs.

Implementation Strategies

Implementation strategies are defined as methods or techniques designed to enhance implementation outcomes (Proctor et al., 2013). Implementation strategies can be conceptualized as the "interventions" that are adult-facing and aim to facilitate behavior change. For example, large group didactic training is considered a cornerstone pre-implementation strategy designed to increase implementer knowledge about a given program or practice to increase the likelihood they will perform behaviors consistent with the program or practice. Another example, coaching, represents an active-implementation strategy designed to ensure proper adoption and sustained implementation of the new practice. Researchers have established three different types of implementation strategies based on their composition: (a) discrete (i.e., a single tactic, process, or action), (b) multi-faceted (i.e., two or more discrete strategies), or (c) blended (i.e., two or more discrete strategies that are packaged as a protocol or a branded multi-component implementation intervention; Proctor et al., 2013).

Although researchers have started to explore differential effects according to the type of implementation strategy (Powell et al., 2014), it is unclear which type of strategy may be most effective to promote individual-level implementation outcomes. Researchers have also established taxonomies of implementation strategies (Powell et al., 2015) to provide a common nomenclature for researchers and practitioners when labeling and describing the strategy that was selected and used to support implementation. Over 70 different implementation strategies have

been identified, and recently this work has been adapted to the school context (Cook et al., 2019). Although some implementation strategies focus on outer setting factors of a school (e.g., policy, funding, state and district leadership) and inner organizational factors of a school building (e.g., leadership, climate, protected time; Powell et al., 2015), other strategies focus more directly on implementer behavior change and range from providing ongoing consultation, delivering dynamic training, and creating checklists and protocols to discrete strategies such as planning, teaching or modeling skills, and providing performance-based feedback (Powell et al., 2015; Proctor et al., 2013). In this study, we focused on individual-level strategies to promote general education teacher implementation.

Characteristics of Implementation Strategies. In addition to the type of implementation strategy, there are several characteristics that are important to capture and examine in relation to the impact they have on implementation outcomes. Capturing the characteristic features of an implementation strategy is important to operationally define, permit future replication, and identify the core features of implementation strategies that promote implementation outcomes. Proctor and colleagues (2013) provided reporting guidelines for characterizing important features of an implementation strategy, including (a) the actor (i.e., who enacts the strategy), (b) action(s) (i.e., what are the specific actions, steps, or processes that need to be enacted), (c) action target (i.e., where and to whom the strategy is delivered), (d) temporality (i.e., when the strategy is used), (e) dose (i.e., intensity), (f) implementation outcome (i.e., implementation outcome(s) likely to be affected by each strategy), and (g) justification (i.e., empirical, theoretical, or pragmatic justification for the choice of implementation strategy).

Behavior Change Techniques and Mechanisms of Action. Contemporary research on implementation strategies has moved beyond simply describing and labeling strategies; current

research instead focuses on specifying the irreducible core components or active ingredients of implementation strategies and understanding how they work and likely influence implementation outcomes, such as behavior change (e.g., Michie et al., 2016). One prominent method includes identifying the behavior change techniques (BCTs), which are the "observable, replicable, and irreducible component(s) of an intervention designed to alter or redirect causal processes that regulate behavior" (Michie et al., 2013 p. 82). Examples of BCTs within an implementation strategy include action planning, emailed prompts, and performance feedback.

In addition to specifying the BCTs involved in an implementation strategy, it is also beneficial to define their targeted mechanisms of action (MoA; i.e., the processes or events through which a BCT operates to obtain its effect; Lewis et al., 2018; Michie et al., 2016). MoAs are informed by behavior change theory and allow for an understanding of the *how* or *why* a given BCT achieves its effect (Michie et al., 2016). Examples of MoAs include beliefs about one's ability to carry out a behavior, beliefs about the outcome of changing a behavior, social influences, and behavioral cueing.

Established guidelines have improved strategy specification; however, many implementation strategies have not yet been specified well enough to be linked to precise MoAs in a coherent manner (Lewis et al., 2018). For example, coaching and consultation represent a term for a generic category of implementation strategies for which the behavior change techniques (BCTs) are underspecified. This ambiguity makes it difficult to identify the precise MoAs that enable coaching and consultation to effect teacher behavior change (Nadeem et al., 2013) and in turn make it challenging to synthesize data across studies and describe effective aspects of coaching and consultation with precision (Lewis et al., 2018; Michie et al., 2009). Coaching and consultation, however, often include multiple BCTs targeting specific MoAs

outlined in the literature, such as dynamic didactics to target knowledge, modeling to target self-efficacy, role playing or directed rehearsal to target skill acquisition, and beliefs about capabilities, or goal setting to target behavioral intention and motivation (Michie et al., 2016).

Michie et al. (2013) utilized a taxonomy of behavior change techniques (BCTs) and their associated MoAs (Carey et al., 2017) were used to identify the number of BCTs and categorize the types of BCTs in each implementation strategy, as well as specify the precise MoAs. This is important to provide a better understanding of *how* the discrete components or active ingredients of implementation strategies work and *why* implementation strategies may exert a desired effect on implementation outcomes (Cane et al., 2012). Michie and colleagues (2013) developed a taxonomy of 93 distinct BCTs, constituting a method for specifying the "active ingredients" of implementation strategies targeting behavior change. Michie and colleagues (2018) have also engaged in extensive research to identify the most common mechanisms from a larger set of behavior change theories. This resulted in the identification of 26 MoAs that can be used to understand the targets of BCTs and provide a more precise understanding about the *how* (BCTs included) and *why* (MoA targeted) implementation strategies work (Cane et al., 2012; Michie et al., 2013, 2018).

Single Case Experimental Designs

Single case experimental designs (SCEDs) employ an idiographic approach to research and utilize small-n samples, which allow for implementation strategies to be tailored to individual changes in behavior over time (Ledford & Gast, 2018). SCED's are defined as a small-n study that involves manipulating an independent variable and repeatedly studying dependent variables over time. A wide variety of SCEDs exist; however, a key component of SCEDs is that experimental control can be demonstrated by repeatedly measuring some target

behavior that is altered or systematically withdrawn across time, within or among participants, across setting, or across varying behaviors (Byiers et al., 2012; Ledford & Gast, 2018; Smith, 2012). At their most basic, SCEDs start with a baseline condition (A) followed by some intervention condition (B) (Gast & Ledford, 2018). To qualify as an SCED, however, experimental control needs to be demonstrated through repeated occasions or replications (e.g., in an A-B-A-B or withdrawal design, baseline data is collected in the first condition [A], followed by introduction of some intervention [B], followed by a withdrawal of the intervention [A], followed again by re-introduction to the intervention [B] to discern whether the effect can be replicated). SCEDs provide benefits such as cost-efficiency and flexibility compared to group-based randomized control trials (RCTs), while still being experimental (Barlow & Nock, 2009); RCTs take a nomothetic approach to research and measure performance a limited number of times and capture large group averages, losing valuable data that may not be representative of any individual (Barlow & Nock, 2009; Krasny-Pacini & Evans, 2018). There are also fundamental concerns among researchers regarding group homogeneity and making inferences to individuals using nomothetic data (Ledford & Gast, 2018; Westen et al., 2004). Given the impact that individual-level characteristics have on implementation outcomes (i.e., attitudes, self-efficacy, outcome expectancies, and intentions; Locke et al., 2019) along with the desire to study behavior change at the individual level (Lobo et al., 2017), as well as their prominence in the literature (Noell et al., 2014), SCED studies were selected for analysis in this study.

Existing Reviews of School-Based Implementation

Rigorous school-based implementation studies targeting individual behavior change date back to the late 1990s (Noell et al., 1997). Since then, there have been four systematic reviews dedicated to examining school-based implementation strategies. Solomon and colleagues (2012)

analyzed the effects of performance feedback on teacher implementation of academic and behavioral interventions across 36 single-case studies, finding moderate overall effects (Improvement Rate Difference [IRD] = .62). Noell and colleagues (2014) used mixed linear models to analyze implementation strategy efficacy for improving teacher treatment plan implementation across 29 single-case studies. These authors focused on general and special education teachers and analyzed the effects of performance feedback alone and with multiple components, such as directed rehearsal and meeting cancelation, and found overall large effects. Fallon et al. (2015) conducted a systematic review to examine whether performance feedback met criteria as an EBP. Based on their review, these authors concluded that performance-based feedback was an implementation-oriented EBP. Last, Stormont et al. (2015) systematically reviewed the effects of coaching strategies to improve teacher implementation of social behavior interventions across 29 studies. Stormont et al. found positive improvements across 86% of all studies.

Although these reviews have provided important insight about the effects of school-based implementation strategies, there remain some notable gaps to be addressed through research. First, prior reviews focused on a narrow subset of implementation strategies. For example, both the Solomon et al. (2012) and Fallon et al. (2015) meta-analyses focused on performance-feedback strategies alone, and the Stormont et al. (2015) study analyzed coaching strategies alone. Furthermore, the Noell et al. (2014) review examined only post-training, follow-up strategies and not strategies used in pre-implementation or maintenance phases. Moreover, none of the extant reviews focused exclusively on general education teachers and with EBPs implemented across a multi-tiered system of support (i.e., universal, targeted, and intensive). Last, none of the prior studies leveraged guidelines and frameworks from the broader field of

implementation to increase classification and characterization of implementation strategies. Although prior studies have identified positive overall efficacy across strategies used, they lack specificity for how and why these strategies have achieved their effects. Specifically, implementation strategies have not been analyzed according to the type (i.e., discrete, multifaceted, or blended), key characteristics (i.e., name, actor, action, action target, and temporality), and BCTs and corresponding hypothesized MoA (i.e., knowledge and skills, motivation, or behavior regulation). It is important for school-based implementation research to tap into the broader implementation science literature to not only identify important knowledge that can be used to better capture and explain the types and effects of implementation strategies but also produce generalizable knowledge that contributes to the multidisciplinary science of implementation (i.e., Aarons et al., 2011; Cane et al., 2012; Powell et al., 2015; Proctor et al., 2013).

Purpose of the Current Study

Considering the current gaps in the literature, the purpose of this study was to conduct a meta-analysis of peer-reviewed SCED studies examining the types and effects of implementation strategies when used to increase general education teachers' adherence to EBPs designed to address student SEB needs. Peer-reviewed studies were selected to ensure internal validity among included studies as well as overall replicability (Schmucker et al., 2017). Although it can ameliorate publication bias, grey literature is often incomplete, the study quality is questionable, and retrieving grey literature systematically is cumbersome and difficult to replicate (Adams et al., 2016; McClain et al., 2021). This study employed established guidelines and frameworks from implementation science to extract data on the type, characteristics, and BCTs used in implementation strategies. In so doing, this study sought not only to provide an update on

findings from previous reviews of implementation strategies in schools, but it also aimed to provide specificity concerning the features and effects of implementation strategies to pinpoint directions for future school-based implementation research. The following three research questions guided this meta-analytic investigation of the single-case literature:

1. What types of implementation strategies have been examined to improve general education teachers' adherence to SEB EBPs?
2. What is the overall effect of employing implementation strategies on improving general education teachers' adherence to the SEB intervention?
3. What characteristics (i.e., actors, actions, action targets, temporality, and the number and specific type of behavior change techniques) moderate the effect on general education teacher adherence to the SEB EBPs?

Hypotheses

We expected the extent literature to indicate that implementation strategies have an overall positive effect on teacher adherence to SEB EBPs, consistent with previous meta-analyses (e.g., Fallon et al., 2015; Noell et al., 2014). We also expected there to be a large number of performance-based feedback and coaching/consultation strategies that target mechanisms of knowledge and skill in delivering SEB EBPs (Fallon et al., 2015; Stormont et al., 2015). Finally, we expected most strategies to be delivered during the pre- and active-implementation phases, as examples of maintenance strategies are limited in the behavior-change literature (Dombrowski et al., 2014; Fjeldsoe et al., 2011).

Method

Retrieval Strategies

A systematic search process was performed through searching electronic databases and

reviewing bibliographies of prior systematic reviews of implementation strategies used in schools. PsycINFO hosted by Ovid Technologies and Academic Search Premier hosted by EBSCO were used to identify literature across the psychology and education fields. The following search terms were developed and included in the two electronic search engines: (teacher* OR educator* OR instructor* OR class*) AND implement* AND (intervention* OR procedur* OR program OR treatment* OR perform*) AND (integrity OR adopt* OR fidelity OR reliability OR adhere* OR accuracy). To attempt to control for study quality, only articles published in peer-reviewed journals were searched (Arumugam et al., 2020).

No dates were specified; studies only had to have been indexed prior to the search, which occurred January 20 and 21, 2018, and articles needed to have an English version available. The following existing systematic reviews were identified during the database searches: Fallon et al., 2015; Noell et al., 2014; Solomon et al., 2012; and Stormont et al., 2015. The first author conducted a backward citation search by extracting additional studies from these existing reviews that were included and analyzed in prior meta-analyses and systematic reviews that were not captured in the database search.

The study inclusion flow-chart is displayed in Figure 1. The first and fourth author initially screened titles and abstracts collected through the electronic database search for inclusion based on five key eligibility criteria. Studies needed to (a) use a single case experimental design, (b) occur in a school during regular school hours, (c) involve general education teachers as the primary participants who were (d) implementing any evidence-based SEB prevention or intervention practice, and (e) the dependent variable needed to be adherence to that SEB EBP (i.e., eligibility criteria are described in full detail below). To calculate inter-coder agreement, percent agreement and Cohen’s Kappa were used. Agreement was calculated at

two points during the initial database screening process of titles and abstracts, which occurred between January and August 2018. Coders collected reliability at the beginning of the screening process over the first 150 titles and abstracts and again midway through screening over an additional 100 titles and abstracts. We set our threshold for appropriate percent agreement at 95% (Liddy et al., 2011). If percent agreement cutoffs were not met, codebook refinement was indicated (Belur et al., 2021). Percent agreement of initial screening was 96% and Cohen's *Kappa* was .54 over 4% (n = 250) of the total records after duplicates were removed (N = 6,355). If there was not enough information to determine eligibility during initial title and abstract screening, articles were included for full-text review.

At this time, the first author also conducted a backward-citation search by reviewing bibliographies of existing meta-analyses and identified 18 additional articles for full-text review. Next, full-text review was completed of 167 articles. The first author reviewed all included articles, and the seventh and eight authors conducted reliability across 19% (n = 32) of full-text review articles. Inter-coder agreement was as follows: percent agreement = 97% and *Kappa* = .88. All disagreements were reviewed together by the coders and consensus was reached on whether to include the study. If agreement could not be met, another author was brought in for discussion until a consensus was reached and codebook refinements then were made. Although there are no agreed-upon cutoffs for coverage of inter-coder agreement at various stages of the systematic review process, our agreement results support the claim that our established inclusion criteria were transparent and explicit (Belur et al., 2021).

Eligibility Criteria

This synthesis included single case experimental design studies that used discrete, multifaceted, or bundled implementation strategies to increase teacher adherence to universal,

targeted, or intensive evidence-based SEB practices delivered to children from pre-K to 12th grade. SEB practice was defined as addressing social aspects of performance, specific classroom behaviors (i.e., on-task, compliance, and disruption), emotion regulation, or competence. Therefore, academic interventions such as math, reading or writing, physical education, health and/or nutrition, motor skill development, sexual health programs including HIV prevention, and drug abuse prevention or intervention programs were not included. Only SEB programs were included because significant differences have been found between teachers' implementation of SEB interventions as compared to academic interventions (Solomon et al., 2012). Moreover, the practice needed to occur during regular school hours; after school programs were excluded. Studies that included implementation strategies for SEB interventions and other types of interventions not identified as SEB were included if data were collected and could be disaggregated on teacher adherence of SEB interventions.

Participants and Setting

To be included, studies needed to target general education teachers as recipients of an implementation strategy. Because general education teachers implement EBPs at significantly different rates than special education teachers when given similar supports (Solomon et al., 2012), we focused solely on implementation strategies targeting general education teachers (e.g. classroom teachers in pre-K through elementary) or content-area teachers such as in middle school or high school. Thus, studies whose participants were special education teachers, assistants, or paraprofessionals were excluded. Studies that included mixed-group implementation teams (e.g., general education teachers, aides, special education teachers) were only included if data could be disaggregated. Studies were included if they were carried out in pre-K, kindergarten, primary, or secondary schools, and took place in the U.S. The decision to

include only studies carried out in the U.S. was made to control for potential confounds due to the laws and policies unique to the U.S. that mandate implementation of EBPs to improve SEB outcomes that may be a source of variation impacting the nature of implementing programs in schools (e.g., Every Students Succeeds Act, 2015).

Study Design

SCED studies that aimed to improve adherence to an evidence-based SEB practice were included. In this study, an SEB EBP was defined as any non-academic intervention that involve any of the following: (a) how students interact with others and impacts the quality of their relationships (i.e., social), (b) the difficulties students have regulating and managing their feelings (i.e., emotions), and (c) the behaviors students exhibit that are disruptive to learning environments (i.e., behavior; National Scientific Council on the Developing Child, 2004). A design was considered experimental if it allowed for conclusions regarding a causal relationship between the independent variable and a change in the dependent variable. To determine a causal relationship, study methodology needed to adhere to WWC standards for study quality (Kratochwill et al., 2010). Studies needed to either fully meet standards or meet standards with reservations to be included in the study. The standards were as follows: (a) systematic manipulation of the independent variable, (b) reliable interrater agreement across at least 20% of observations, (c) at least three attempts to demonstrate the effect, and (d) appropriate number of data points per phase (specific to the design). Quality indicators were coded and are discussed in detail below. Designs could include multiple baseline, alternating treatments, or reversal or withdrawal (ABAB) designs (Ledford & Gast, 2018, for more information and examples regarding single-case research).

Dependent Variables

Studies were included if the outcome measure involved adherence to an SEB EBP, which entails determining the proportion of steps, components, or activities of the EBP delivered as intended. Studies collecting EBP adherence data were only included if the collection method was via systematic direct observation or permanent product review. Systematic direct observation involves observing and rating whether the target implemented the intervention steps (Sanetti et al., 2013). Permanent product review involves collecting and analyzing the physical products resulting from an intervention as evidence of adherence (Sanetti & Collier-Meek, 2013). Although systematic direct observation has been shown to be the most accurate method for measuring adherence (Sanetti et al., 2013), both methods have some empirical evidence supporting their use as measures of treatment integrity (Fiske, 2008; Noell, 2008). Studies that relied on self-report were excluded, as prior research comparing self-reported treatment integrity data from teachers to permanent product review and direct observation conducted by trained professionals determined that self-reported data were often overestimated and with reduced variability (McKenna et al., 2014; Noell et al., 2005). For purposes of calculating effect size estimates, studies needed to include graphic, tabular data. The standardized method for collecting and analyzing adherence data is further delineated in the "Data Analysis Plan" section.

The eligibility criteria described above established the universe of generalization for this study. The results of this analysis may be generalized to schools and teachers similar to those represented in the sample, with students who are similar to those represented in the sample (i.e., students attending public schools in the U. S.), and with studies based on the assessments and strategy components represented in the sample (i.e., studies that used an implementation strategy to increase teacher adherence to an SEB EBP as indexed via directly observing or reviewing permanent products).

Coding Studies

Studies meeting inclusion criteria were coded on a variety of study attributes aligned with the proposed research questions. The coding scheme was developed by the first, second, and third authors through an iterative process and included five broad coding categories: (a) setting and participant characteristics, (b) specific implementation strategy characteristics, (c) characteristics of the SEB EBP that teachers were implementing, (d) adherence data collection methodology, and (e) quantitative data extraction. The first and second authors collected inter-coder agreement for (a), (c), (d), and (e), and the first and third authors collected agreement on (b). To promote clarity and high reliability of the coding scheme, the first and second author double coded a random sample of 10 included studies (30%) and met on multiple occasions to compare and refine the definitions of the codes used. The average percent agreement across all categorical variables coded in the study was 95.7% and Cohen's Kappa was .87. In the case that agreement was less than 80% for any code, the two coders met to discuss, refine codes to ensure clarity, and come to a consensus decision. When an agreement could not be met, another author was brought in to consult and reach consensus. One common area of disagreement was on implementation strategy type. Below is a description of each of these coding categories and the specific variables that were extracted from the included studies.

Setting and Participant Characteristics

This category included the author, setting, and participant information, as well as descriptive information about teachers, students, and actors (i.e., those who employed the implementation strategy with teachers). School-wide information included type of school (i.e., preschool/Head Start, kindergarten, elementary, middle, or high school), ethnicity (i.e., % non-white), urbanicity (i.e., urban, suburban, or rural), geographic region, and socioeconomic status

(i.e., the percent of students receiving free or reduced-price lunch). Teacher participant information included total years of teaching experience, ethnicity, and sex. Actors were coded as either being an outside consultant (i.e., researcher or graduate student) or as an internal consultant (i.e., school-based staff member).

Implementation Strategy Specification

The first and third authors completed coding of the specific actions (i.e., name and BCT of each implementation strategy), which were derived from each included article and were used to categorize and code the MoAs according to Michie and colleagues' (2013) taxonomy. Through an iterative process, authors identified the BCTs by analyzing descriptions of implementation strategies provided in study methods and then categorized them under one or more of the 26 MoAs. Mechanism definitions came from the Theoretical Domains Framework Coding Manual (Cane et al., 2012) and an interactive Theory and Techniques Tool that allows users to examine evidence of links between extant BCTs and MoAs within literature synthesis, consensus, and triangulation studies (Johnston et al., 2018). Authors 1 and 3 only coded the treatment phase of each study to assess each strategy's relative efficacy and only coded MoAs that had established links with BCTs. We utilized an established ontology from the Human Behavior Change Project (https://theoryandtechniquetool.humanbehaviourchange.org/) that establishes a link using a criterion of a relationship exceeding the $p < .05$ value in their extensive literature synthesis and ≥ 80% of experts having rated the link as "definitely" in their series of consensus and triangulation studies (Carey et al., 2018; Connell Bohlen et al., 2018). For further information regarding the synthesis of links between BCTs and MoA's, see Carey et al. (2018) and Michie et al. (2021). Authors collected inter-coder agreement of BCTs and MoAs across a random sample of 20% of included studies (percent agreement = 89%, Kappa = .77).

Following Proctor et al.'s (2013) reporting standards, implementation strategies were also categorized by type as either discrete (i.e., one tactic, process, or action), multi-faceted, (i.e., two or more discrete strategies), or blended (i.e., two or more discrete strategies that are packaged as a protocol or branded implementation intervention) and by temporality (pre-implementation, active implementation, or maintenance). Specifically, pre-implementation strategies targeted factors occurring before implementation of the practice takes place (e.g., training or planning), active implementation strategies targeted factors occurring during implementation (e.g., coaching, performance feedback, test-driving), and maintenance strategies were targeted at maintaining adherence over time (e.g., systematic fading).

EBP Characteristics

The name of the SEB EBP being implemented by the teacher and the tier at which the SEB practice was implemented were coded. No definitive source yet exists to guide categorization of interventions tiers beyond general guidelines that have emerged through loose consensus from practitioners and researchers. As such, the tier of support was determined using guidelines from McIntosh and Goodman (2016) on integrated multi-tiered systems of support. SEB EBPs delivered at the class-wide level to all students were considered Tier 1 (e.g., increasing the use of specific praise statements throughout the classroom or implementing the Good Behavior Game), brief individual or small group interventions requiring few resources (e.g., training, time, materials, personnel) for students with identified SEB needs were considered Tier 2 (e.g., Check-In/Check-Out as a brief, individual intervention), and more resource intensive, individualized interventions were considered Tier 3 (e.g., conducting an experimental functional analysis or implementing components of an individualized student support plan). These codes enabled descriptive analysis of the types of EBPs general education teachers were

implementing, as well as an inferential test whether the effects of implementation strategies varied as a function of the tier of support. Finally, social validity data, such as acceptability, appropriateness, and feasibility of the implementation strategies were collected and coded.

Methodology and Quantitative Data Extraction

The method of collecting adherence data (i.e., direct observation or permanent product) and whether maintenance of adherence was collected via a follow-up phase were coded. Teacher adherence to the EBP (i.e., treatment integrity) was determined by tabulating the time-series data. This was accomplished by extracting images of the time-series graphs from PDFs and uploading them into a crosstabulation software, WebPlotDigitizer (Rohatgi, 2019), which has evidence of validity and reliability (e.g., r = .99; Drevon et al., 2017). This software allows users to produce X-Y coordinates from time-series data on a standardized metric to compare effect sizes across studies using different measures (e.g., percentage intervention step adherence, number of praise statements). Coordinates and data points overlap the time-series images and each data point correlates with a Y-value, which can then subsequently be used to calculate various effect sizes. Since each datapoint is represented by a standardized point along the same Y-scale, this allows users to compare effect sizes across studies. The first and second author completed the digitizing and calculated reliability across 34 time-series graphs (22%) for both Cohen's *d* and Tau. Although evidence of reliability and validity was collected using WebPlotDigitizer, there was still a chance for slight differences in how the first and second author completed the process that might result in slightly discrepant effect size estimates. Given that Cohen's *d* and Tau are continuous metrics, we could not calculate reliability as a percent of agreement or Cohen's Kappa. Instead, we established bounds within which estimates needed to be relative to one another (+/- .05). Coders had perfect or near perfect agreement (within a

difference of .05) across 74% of effect sizes and the overall average difference of effect sizes between coders was minimal (d = .09). In cases of disagreements greater than .05, coders reviewed the data together and re-calculated the effect sizes until differences were within the bounds of perfect or near perfect agreement ($d \leq .05$). Additionally, a two-tailed t-test was calculated between the two sets of effect sizes to further support evidence for reliability, which indicated a nonsignificant difference (p = .95). Authors determined whether the effect was maintained (coded binarily as either *yes* or *no*) through visual analysis of data between adjacent phases (Lane & Gast, 2014). The authors were unable to locate a method for establishing whether an effect was maintained in the literature, therefore an effect was determined to be maintained if data in the maintenance phase represented more than 50% overlap as compared to the adjacent treatment phases. Data that did not overlap in the maintenance phase and were in the negative direction, which indicated a deleterious effect, were coded *no*.

Quality Indicators

In order to determine methodological quality among the included SCED studies, three components have been shown to be appropriate to establish an indication of study quality within the current literature base of SCED studies and among the treatment integrity literature (Lobo et al., 2017; Sanetti & Kratochwill, 2009a): (a) because assessment of the dependent variables involves raters of teacher implementation behavior, inter-rater reliability of adherence data collection was selected (Reichow et al., 2008); (b) because knowing whether the implementation strategy was delivered to the teacher is needed to indicate that the teacher actually received the implementation support, fidelity of strategy implementation (i.e., procedural reliability) was collected; and (c) methodological design standards within SCED were used to determine experimental control (Kratochwill et al., 2010; Lobo et al., 2017). Because no index validated for

the specific use of assessing methodological quality of implementation strategy efficacy using a SCED was available, a summative scale of study quality was developed for use in this study. The scale was intended to capture overall methodological rigor of SCED combined with important implementation factors. The contents of this scale are detailed next.

First, all SCED studies were evaluated by the first author using methodological design standards outlined by the What Works Clearinghouse to assess internal validity of results from SCED research. The first author familiarized himself with the standards through studying the manual (Kratochwill et al., 2010). Accordingly, studies were coded as either meeting standards or meeting standards with reservations, and they were scored as "2" and "1" respectively. Studies not meeting the standards were excluded from the review.

Second, the scale included inter-rater agreement among the actors (i.e., those delivering the implementation strategy to the teachers). Because the data collection method of the dependent variable (i.e., direct observation or permanent product) is a key component of study quality, we coded inter-rater agreement among studies. However, no empirical guidelines regarding the level of fidelity of implementation that is needed to realize a meaningful gain in student performance exists (Schulte et al., 2009). Therefore, we developed the following cut-off points based on recommendations by Bergman (2018): (a) if reliability was above 90%, studies were scored a "2"; (b) if reliability was between 80% and 89.4%, studies were scored a "1"; and (c) if reliability data were not collected or reliability was below 80%, it was coded "0".

Finally, the first and second authors coded for whether fidelity of strategy implementation (i.e., whether the proposed implementation strategy was delivered by the actor to the teachers with fidelity) was collected. If studies collected these data, they were scored a "1", and if not, they were scored a "0". Combining these three metrics comprised the overall study

quality scale for included studies, which was rated on a scale of 1–5. Studies receiving a higher score indicate stronger evidence of methodological quality.

Data Analysis Plan

A funnel plot was used to examine publication bias in the results. A funnel plot is a scatterplot of effect sizes (x-axis) from individual studies compared to a study's precision (y-axis), in this case, standard error. Studies with larger effect sizes are placed toward the top and studies with smaller effect sizes should scatter more widely at the bottom of the plot, thus creating an inverted funnel shape (Sterne et al., 2011). If studies are underrepresented in the bottom-left quadrant of the plot, this could indicate publication bias, as studies with smaller effect sizes are less likely to be published. To further test for plot symmetry, Egger's regression test was completed, which determines whether there is a relationship between the observed outcomes and the chosen predictors. A significant relationship would indicate further evidence of publication bias (Egger et al., 1997).

A fail-safe N for effect sizes (Orwin, 1983) was computed in *R* to determine the potential influence of a file-drawer problem (Rosenthal, 1979). According to Cohen's (1988) guidelines, a phi coefficient less than 0.30 is considered negligible, 0.30–0.49 is small, 0.50–0.69 is moderate, and 0.70 or above is strong. The criterion phi for a negligible effect of 0.29 was used via the *fsn function* of the *metafor package* in R. This metric determines the number of null effect sizes (phi = 0) needed to bring the average effect size (phi) below 0.30.

Handling Missing Data and Outliers

The authors anticipated missing data on some study characteristics. Variables were not included as moderators if, across all studies, less than 75% of studies reported data for that variable; however, descriptive statistics were still calculated and reported for these variables. No

established guidelines to make decisions about percent missingness currently exist, so we decided to set a cut point of 25% based on recommendations (van Buuren, 2018). To examine the maximum number of moderators without biasing the estimates by leaving out effect sizes, multiple imputation methods common among meta-analyses were used via the MICE package in *R* (van Buuren & Groothuis-Oudshoorn, 2011). Whereas continuous values can be imputed, they carry the assumption that data is missing at random, which cannot be tested, and calls for caution in the interpretation of variables with missing data. In the case that an effect size was outside 3 SD of the mean in either direction, those variables were treated with a Winsorizing method; effect size values were placed at the exact value of furthest outlying variable within 3 SD of the mean in the appropriate direction (Lipsey & Wilson, 2001).

Effect Size Metrics

When analyzing SCED studies, it is recommended that both parametric (regression-based) and nonparametric (non-overlap) effect sizes are collected (Kratochwill, et al., 2010). To answer Research Question 2, we calculated three SCED effect size metrics—Tau, Tau-U, and a small-sample corrected *d*-statistic (i.e., Hedge's *g*) developed for use with SCEDs (Hedges et al., 2012). Tau is a method for measuring nonoverlap, carries non-parametric assumptions, and is derived from the Kendell Rank Correlation and the Mann-Whitney *U*-test (Parker et al., 2011). Tau was selected because it uniquely allows for a monotonic baseline trend correction (i.e., Tau-U; Parker et al., 2011), which has been cited as a crucial limitation of other single case effect size metrics (Parker et al., 2011). Moreover, Tau combines the strength of another powerful nonparametric effect size method (i.e., nonoverlap of all pairs [NAP]; Parker & Vannest, 2009) to calculate all pairwise data comparisons with the option for trend correction if monotonic baseline trend is found via significance testing. When interpreting Tau and Tau-U, a .20 value is

considered a small change, .60 a moderate change, .80 a large change, and above .80 a large to very large change (Vannest & Ninci, 2015). Tau statistics and trend analysis were conducted using calculators at singlecaseresearch.org (Vannest et al., 2016).

A *d* statistic developed for SCD (i.e., Hedge's *g*) was calculated in R following guidelines from Shadish et al. (2014). The values produced by *g* represent standard deviations and should be interpreted as such (Shadish et al., 2014). Because *g* carries assumptions of normality and heterogeneity of variances, model checking procedures were conducted via qq-plot and a radial plot. Regression-based quantitative methods, although controversial among SCED scholars due to potential assumption violations (Kratochwill et al., 2010), were used to allow for meta-regressive methods to be used when analyzing within and between study heterogeneity in effect size (Hedges et al., 2010). To determine maintenance effects of studies that collected follow-up data after supports were removed, the first author conducted visual analysis and determined whether treatment levels were maintained after supports were removed and coded as either *yes* or *no*, with the cutoff being greater than 50% of data-points still overlapping with the treatment phase to be considered as evidence supporting maintenance of effect.

Heterogeneity and Moderator Analysis

Heterogeneity in effect size estimates were examined both descriptively and inferentially. Descriptive comparisons were conducted when there was an insufficient number of studies to conduct inferential moderator analyses with adequate power. A random effects model was created via the *robumeta* and *metafor* packages in *R* (Fisher & Tipton, 2015; Viechtbauer, 2010). The *Q* test, τ^2, and I^2, were used to assess heterogeneity. The *Q* test (Kulinskaya & Dollinger, 2015) is used to assess whether there is true heterogeneity between studies. This test is conducted

by summing the squared deviations of each study's effect estimate from the overall effect estimate, while weighting the contribution of each study by its inverse variance. In short, each teacher participant produces a unique effect size, and each study produces an overall weighted effect. By using the inverse variance, studies that produce a large amount of effect size variability contribute less to the overall average effect (i.e., teachers within the same study having vastly different experiences will impact the overall average effect size less). τ^2 represents the between study variance component in the correlated effects meta-regression model and the between-cluster variance component in the hierarchical effects model. I^2 provides information about whether the spread of effect sizes is due to sampling error and quantifies the amount of variability in effect size estimates due to effect size heterogeneity as opposed to random variation (Lipsey & Wilson, 2001). If the *Q* test was significant and the between-study variance had a value of ($I^2 > 25\%$) an omnibus moderator analysis was conducted.

The following variables were included in the moderator analysis: (a) the grade taught by the teacher, (b) the tier of support at which the SEB EBP was delivered by the teacher (i.e., tier), (c) the number of BCTs included in the implementation strategy to influence teacher adherence, (d) study quality, and (e) the number of years of teaching experience. Rationale for including each variable as a moderator is provided next. Teacher training in basic behavior change across different grade levels, particularly between preschool and elementary, and middle school, emphasizes different strategies teachers may use based on the instructional and developmental needs of their students (e.g., younger children needing shorter, explicit behavior guidance directives than teens). As such, teachers trained for instruction at different levels may vary in their familiarity with some of the techniques used to support their own implementation efforts, which could impact their effectiveness. Similarly, years of experience was included to determine

whether prior experiences teaching impacted response to implementation strategies. Less experienced teachers may be more open to new practices and in need of support; however, teachers with more experience may be more skilled in their delivery of effective practices but may also be more resistant to change. The level of complexity of the EBP being implemented by teachers, measured by EBP tier (e.g., Tier 1 less intensive, Tier 3 requiring more intensive efforts), was hypothesized to moderate fidelity of implementation. A quality indicator was included in the moderator analysis to determine if study quality impacted results. Finally, the number of BCTs within a given implementation strategy or package should be associated with higher rates of implementation. Hypothetically, a greater number of BCTs should act upon a wider variety of mechanisms of action and be associated with larger effect sizes. Post-hoc analyses were conducted if results indicated significant moderators.

Results

Study, Participant, and Setting Characteristics

After duplicates were removed, the initial database search yielded 6,355 articles, and after screening titles and abstracts, 149 studies were included for full-text review via database search. An additional 18 articles were identified through bibliographic searches of articles included via database search, as well as studies included in past reviews, bringing the total number of articles to 167. Full text review yielded 28 studies for inclusion. Studies included in this analysis were published between 1992–2017, and 122 unique time series graphs provided effect sizes for the meta-analysis. Study and participant characteristics are displayed in Table 1. Each study had an average of 3.4 participating teachers (min = 1, max = 6). Of the studies that reported teacher participant characteristics, the majority were female (99%) and White (84%) and had an average of 11 years of teaching experience. Most studies took place in elementary

schools (60%) in the Northeast region of the U.S. (48%), and in suburban areas (56%). No studies included high school students. Many studies did not report on key participant and setting characteristics; for example, only 32% of studies provided school-wide ethnicity, 52% reported urbanicity, and 29% reported a measure of socio-economic status. Only 19 (61%) studies reported procedural reliability of the delivery of the implementation strategy and 16 (52%) collected quantitative social validity evidence.

Study Quality and Methodology

Of the 28 studies included in this review, all met WWC experimental standards for single case research (14 met full standards and 14 met standards with reservations; Kratochwill et al., 2010). Overall, study methodological quality was high, with a mean score of 4 (range: 3–5) on a scale of 1–5, with 5 being the highest possible quality index score and 1 being the lowest. Five studies obtained a score of '5', 18 studies obtained a score of '4', and five studies obtained scores of '3'. Most studies (n = 24) collected teacher adherence via direct observation and four studies used permanent product review. All studies used a multiple baseline across participants design.

Research Question 1: Types and Characteristics of Implementation Strategies

In total, 15 different named implementation strategies were used across studies to promote general education teachers' adherence to SEB EBPs, which are displayed in Table 2. Across the 28 studies, 17 (61%) used a multifaceted strategy, 9 (32%) used a blended strategy, and 2 (7%) used discrete strategies. Results associated with coding and analyzing distinct BCTs revealed that implementation strategies included 23 unique BCTs (Table 3) across 16 MoAs (Table 4). On average, studies used 4 BCTs, with a maximum of 9 and a minimum of 1. Performance feedback, providing procedural knowledge of intervention steps, and action

planning were the most researched BCTs, and strategies such as skill generalization, self-evaluation, and collaborative team working were among the least researched. Regarding MoAs, 19 out of the 26 (73%) were used at least once while coding. The average study included implementation strategies that targeted 7 MoAs with a maximum of 13 and a minimum of 2. The most common MoAs were Feedback Processes, Beliefs about Capabilities, Skill, and Motivation.

Most actors (i.e., those who delivered implementation strategies to the teachers) were outside experts, such as PhD-level researchers, graduate students, or consultants (n = 25; 89%), with only three studies (11%) using employed school-based staff (i.e., administrators, instructional coaches, social-workers, or special education teachers) to deliver implementation strategies. Among the 122 time-series graphs, 44 (33%) included a pre-implementation strategy, 119 (91%) included an active-implementation strategy, and 11 (8%) included a maintenance strategy. Out of the 122 time-series graphs, follow-up data at one or two months after study completion were collected for 49 (37%), and of those, 33 (63%) showed a maintained effect once supports were removed. When examining data regarding the tier or level of support, 15 (54%) aimed to increase teachers' delivery of a universal Tier 1 EBP, 8 (28%) were used to increase teacher adherence to a targeted Tier 2 intervention, and 5 (18%) were deployed to support teacher adherence to an intensive Tier 3 intervention.

Research Question 2: Magnitude of Effect

After checking assumptions for normality, no major violations were found. The process of extracting time-series data yielded minimal outliers (6% of participants). Therefore, Winsorizing methods of the standardized mean difference effect sizes were used to restrict outliers to the most extreme data point within 3 SD from the mean effect size across all 122 time-series graphs. This process lowered the average effect size (from g = 2.51 to g = 2.32) and had

minor effects on the average I^2 value (from $I^2 = 87.2$ to $I^2 = 82.63$). Eight outlier data points (6%) were found and windsorized, resulting in the following overall effect sizes: Hedge's $g = 2.32$ 95% CI [1.88, 2.76], Tau = 0.77, 95% CI [0.66–0.88]. Baseline trend was found in 13 (9%) time-series graphs and was corrected when calculating Tau estimates. All adjusted Tau (i.e., Tau-U) values were similarly windsorized to account for potential inflation during trend correction. Average effect sizes for each study are displayed in Figure 3 and Figure 4, and the number and average effect sizes for each BCT and MoA are presented in Table 3 and Table 4.

Publication Bias

Figure 2 presents the funnel plot for publication bias. Data clustered toward the top of the funnel around the average effect size, indicating that most studies had high precision. When studies had low precision, however, they tended to fall in the bottom-right quadrant, indicating higher effects. There was a lack of data points represented in the bottom-left quadrant (studies of low effect size and low precision), which provides evidence that publication bias may have impacted results. This analysis is corroborated by results of Egger's test, which were significant ($z = 11.384$, $p < .0001$), further indicating asymmetry.

A fail-safe N for effect sizes (Orwin, 1983) was computed to determine the potential influence of a file-drawer problem (Rosenthal, 1979). Results indicated that an additional 138 null effect sizes would be needed to diminish the effects found in this study. As each study had an average of 3.3 participants, this would mean another 42 studies of all null effects would be needed, which is nearly double the number of studies included in this analysis.

Effect Sizes by Type, Behavior Change Techniques, and Mechanisms

Although there was no statistical difference between strategy type, average effect sizes of multifaceted strategies were associated with the highest effect size ($g = 2.60$), followed by

blended strategies (g = 2.27), and discrete strategies (g = 2.18). All BCTs could be interpreted as being associated with "moderate" to "large" average effect sizes (Table 3) with minimal variation. Across BCTs present in at least five studies, identifying and problem-solving barriers to implementation were associated with the smallest relative effect size (g = 1.99, Tau = .72) and video modeling was among the largest relative effect sizes (g = 3.70, Tau = .97). When examining effect sizes according to the MoAs, Motivation, Beliefs about Consequences, and Behavioral Regulation were associated with the smallest effect sizes, and Social Learning and Imitation and Social Influences were associated with the largest effect sizes (Table 4).

Research Question 3: Moderator Analysis

There was a significant amount of between-study heterogeneity within the sample (I^2 = 82.63, τ^2 = 1.63, Q(df = 121) = 555.23, $p < .001$), lending itself to moderator analysis. The following moderators were included in the omnibus model: (a) the number of BCTs used, (b) the intervention tier in which the SEB EBP was delivered, (c) the grade each teacher taught, (d) study quality indicator, and (e) the total years of experience (Table 6). In this analysis, only the total number of BCTs used was significant (z = 3.80, p = .0001). The standardized estimate was positive (.49), indicating that the more BCTs used within an implementation strategy were associated with larger effects. An exploratory post-hoc analysis was conducted to further understand the relationship between the number of BCTs incorporated and effect size. Specifically, we were interested in determining if the relationship was linear or curvilinear (i.e., if there was a diminished return in the number of BCTs used relative to the overall effect). Therefore, we created a linear and curvilinear model (taking the square root of the number of BCTs), and results indicated that a curvilinear model explained slightly more variance (R^2 = 19.92%) as compared to a linear model with (R^2 = 18.56%). Figure 5 demonstrates this

curvilinear relationship, showing the largest increase in average effect size occurring between 4 and 6 BCTs, with diminished returns beyond incorporating 6 BCTs.

Discussion

The development and strategic testing of implementation strategies designed to increase general education teachers' delivery of EBPs represents a growing body of research that aims to address the science-to-practice gap in schools (Atkins et al., 2016). Implementation strategies used to promote general education teachers' EBP adherence vary considerably by their type, complexity, and the specific BCTs included that target various MoAs. We categorized the extant research examining the use of implementation strategies to promote teachers' adherence to EBPs targeting student SEB outcomes in accordance with guidelines and frameworks drawn from the implementation science literature (Proctor et al., 2013). These included strategy type (i.e., discrete, multi-faceted, and blended), temporal stage (i.e., pre-implementation, active implementation, and maintenance), the actor (i.e., who delivered the implementation strategy), and the actions involved (i.e., the name of the implementation strategy and number and type of BCTs used as well as their associated MoA). Additionally, we examined effect sizes associated with various strategy characteristics. Altogether, this work is intended to provide an updated review of the literature to provide concrete recommendations to teachers, administrators, consultants and coaches, implementation intermediaries, and other professionals working to strengthen individual implementation of evidence-based SEB interventions in schools.

As predicted, overall, findings indicated that implementation strategies increased general education teachers' adherence to SEB EBPs. Consistent with prior reviews, results indicated that strategies were associated with moderate to large effect sizes overall (e.g., Fallon et al., 2015; Noell et al., 2014). This is an important finding given the established link between fidelity and

student outcomes (Durlak & DuPre, 2008). As expected, implementation strategies targeting individual-level determinants (i.e., adherence) were most often delivered during the active implementation stage and most frequently involved the use of performance-based feedback. Results from this study indicate that there is a need for school-based implementation researchers to continue developing innovative implementation strategies targeting a wider variety of MoAs in all phases of implementation—specifically, the pre-implementation and maintenance phases. Recent studies have begun to investigate pre-implementation strategies grounded in motivational interviewing principles and behavior change theory (Larson et al., 2020; Lyon et al., 2019).

Little is known about the efficacy of implementation strategies to support teacher sustainment when active implementation strategies are withdrawn. A recent systematic review has shown that sustainability relies on schools giving teachers greater autonomy over the practices selected as well as commitment and support from school leaders and retaining knowledgeable, skilled, and motivated informal leaders, such as coaches (Herlitz et al., 2020). This is consistent with the broader implementation science literature, where much less attention has been paid to examining strategies deployed to support practitioners' EBP sustainment (Moullin et al., 2019). Implementation strategies such as policy alignment, ongoing fidelity audits, and feedback as part of a data-driven continuous improvement process and dedicated professional learning communities to reflect on ways of maintaining EBP implementation represent promising strategies to support EBP sustainment (Squires et al., 2014).

This study found a significant difference in effect between strategies incorporating more BCTs, which is congruent with studies linking improved outcomes to strategies acting on multiple mechanisms of behavior change (Michie et al., 2018). As mentioned, a total of 23 unique BCTs were used across the implementation strategies to influence general education

teacher behavior change. This represents a small subset of behavior change techniques relative to prior research that has identified over 70 unique BCTs that can be used to facilitate implementer behavior change (Michie et al., 2013). Some of these BCTs, such as financial incentives, are less feasible in our modern resource-scarce school settings, but others can be modified to fit the school context, such as those targeting organizational contexts (i.e., changing the social environment and providing protected time for planning and teaming) that have direct and indirect influences on individuals.

Linking BCTs and MoAs

The BCTs involved in the included studies were mapped across three broad clusters of mechanisms. First, many BCTs were linked with the mechanisms *Knowledge, Skills,* and *Beliefs about Capabilities*. These mechanisms seek to increase general education teachers' procedural knowledge of how to implement the different steps or practices of an intervention, improve their ability to deliver the intervention skillfully, and to increase implementer beliefs about their capabilities to implement (i.e., self-efficacy). This finding is intuitive because strategies such as group-based trainings are cornerstone implementation strategies dedicated to increasing implementer knowledge and skills related to the implementation of an EBP, while also building confidence and self-efficacy (Beidas & Kendall, 2010).

Another set of BCTs mapped onto *Feedback Processes, Motivation, Intentions,* and *Beliefs about Consequences*, which involved providing performance feedback after observing the teacher in the classroom and reviewing student behavior data. Indeed, performance feedback is one of the most widely studied school-based implementation strategies (e.g., Stormont et al., 2015). Additionally, motivation and beliefs regarding the consequences of behavior have been linked as important components in behavior change theory (i.e., the Theory of Planned Behavior,

Ajzen, 1991; health action process approach, Schwarzer, 2008).

A third broad cluster of BCTs mapped onto *Behavioral Regulation, Behavioral Cueing, Memory, Attention, and Decision Processes, Reinforcement,* and *Environmental Context and Resources*. These strategies included action and coping planning (i.e., laying out when, where, and how one will implement the components of an EBP as well as planning for potential barriers and relapse), problem solving contextual barriers, adapting components of the intervention to better fit their student population or classroom environment, and providing prompts to help cue teachers to remember to implement. These strategies generally sought to increase adoption and adherence through either changing the environment or how participants or practices worked within their environment. Mechanisms in this cluster identified the when, where, and how of behavior, and are theorized to be effective with people who are properly motivated to implement a change (Schwarzer, 2008).

Moderator Analysis

The moderator analysis revealed a few noteworthy findings. A greater quantity of BCTs/MoAs used to promote teacher adherence was associated with a larger magnitude of effect. This aligns with previous research showing that blended or multi-component implementation strategies may have a larger effect on implementation behavior than discrete, single component strategies (Powell et al., 2019). Moreover, these findings align with prior work indicating that BCTs acting on multiple MoAs impart a larger effect on behavior change (Michie et al., 2016, 2018). Our findings similarly indicate that including too few BCTs may only target a limited number of MoAs and lower the likelihood that adherence will improve. Additionally, post-hoc analyses indicated that using a “kitchen sink” approach and incorporating more than six BCTs in a given implementation strategy was met with diminished returns. Implementation

scientists have encouraged the development of efficient and effective implementation strategies that include a parsimonious amount of theoretically informed behavior change strategies and advocate for a theory-based tailoring approach to strategy development and selection (Lewis et al., 2018). Given that approximately two-thirds of EBP implementation efforts end in failure (Damschroder et al., 2009), it is crucial to uncover the driving factors that determine success. When considering the findings in this review, it is incumbent upon future researchers to scrutinize how implementation strategies are packaged by understanding the specific BCTs included in them, intentionally matching them to the MoAs that fit best with the needs of their target audience and sequencing them in line with best evidence. Doing so will increase the likelihood that implementation efforts will result in successful adoption and sustained use of EBPs.

Limitations and Future Research Directions

This study has several limitations that pinpoint directions for future research. First, applying meta-analytic techniques to SCED data involving multi-phase time-series data to examine strategy efficacy is limited because only adjacent phases can be compared (Lane & Gast, 2014). Furthermore, a notable limitation was the necessity to include studies that involved baseline phases with no support provided (e.g., a "true" baseline) with studies where all participants received cornerstone training practices common across nearly any professional development efforts prior to baseline data collection. Hypothetically, if all participants received this support prior to baseline (with only some participants proceeding through the MBD), then all baselines would be elevated and any "above and beyond" treatment effects for an implementation strategy would be diminished relative to participants that received absolutely no pre-baseline support. This is less an issue in this meta-analysis and instead a valid critique of the

literature. Much of the literature involving SCEDs to test implementation strategies emerge from natural contexts (i.e., coaching or consultation to support a practice). In these natural conditions, teachers receive some form of training, even if it only includes didactics to target knowledge. A true baseline is difficult to establish in real-world contexts, especially for learned behavior. For the most rigorous test of these strategies, studies would need to involve participants with no prior exposure to an EBP and then receive the implementation strategy under investigation. These studies are rare in this literature and arguably lack external validity as compared to studies born from an action research effort involving pre-baseline support. As such, by including those studies, this meta-analysis is likely better able to estimate "real world" effects of these strategies.

It is also important to consider that the effect size metrics (i.e., Hege's *g*, Tau, and Tau-U) used in this study have strengths and limitations. Hedge's *g* follows parametric assumptions although SCED data are non-parametric in nature. However, it allows one to readily rank studies and assess efficacy of maintenance interventions, which Tau is unable to provide as it depends on the calculation of overlapping data. Tau and baseline corrected Tau (Tau-U) have non-parametric assumptions and can be adjusted for baseline trends to capture treatment effects more accurately. However, because available non-parametric effect sizes such as Tau rely on overlapping data, estimates are limited by ceiling effects (Parker et al., 2011). Furthermore, Tau-U is most appropriate when baseline and treatment phases are identical, and estimates can become biased when baseline phases are significantly longer than treatment phases, such as in multiple-baseline designs, which were the most commonly used design across studies in this review (Pustejovsky, 2016). Tau-U was used for only 9% of time series graphs in this review, so potential for estimate bias was minimal. However, researchers should continue to explore different single case effect size estimates that lend themselves to estimating the magnitude of an

effect produced by an independent variable and enable the aggregation of effect sizes across studies with minimal bias (e.g., Shadish et al., 2014).

Further analysis of potential moderators is warranted given the significant amount of between study heterogeneity left to be explained. Teacher-level factors, such as intentions to implement, stress, and burnout represent promising avenues to address potential moderators that mitigate the impact of implementation strategies on general education teacher EBP adherence (Larson et al., 2018; Lyon et al., 2019). Future research should also explore potential moderators, such as teacher perceptions of the feasibility, acceptability, and appropriateness of strategies to better understand with whom implementation strategies work, as well as organizational-level moderators, such as implementation leadership, climate, and citizenship to help determine under what conditions implementation strategies work.

Only articles appearing in peer-reviewed journals were included, and evidence from the funnel plot and Egger's test indicated asymmetry, which is problematic because published studies tend to have a positive bias (Rothstein et al., 2005). Not including grey literature in this study excludes conference proceedings, dissertations, and unpublished studies, for example. To abate this, a fail-safe N was calculated, which indicated that an additional 42 null studies would have been needed to extinguish the positive effects found. Additionally, in one study that was excluded, the authors stated that one participant who did not respond to implementation supports was intentionally removed from the publication. For implementation research, this is a serious concern as it fails to provide a full picture of the findings of an implementation strategy and contributes to the ongoing replication crisis (Makel & Plucker, 2014). It has the potential to further bias the published literature through conscious selection of participants that respond. Although future meta-analyses may also include studies from the grey literature to ameliorate the

publication bias found in this review, other research has indicated bias may exist in grey literature as well (Adams et al., 2016) and that the impact grey literature has on meta-analyses is unclear (e.g., Schmucker et al., 2017). Nonetheless, there is a need to continuously update meta-analyses to allow for the inclusion of unpublished high-quality grey literature.

Another limitation arose during attempts to code implementation strategies with specificity. There were a variety of BCTs incorporated within implementation strategies such as "coaching", "consultation", and "direct training." For example, BCTs such as directed rehearsal, modeling, guided practice, and positive reinforcement were used within these overarching strategies (e.g., Dufrene et al., 2012). In some cases, study authors provided insufficient strategy description to determine the BCTs included in an implementation strategy, limiting efforts to provide specification of the underlying MoAs. To abate this issue, multiple coders inferred the associated mechanisms and collected reliability data. However, the field of implementation science would benefit from increased specificity in describing the BCTs incorporated into implementation strategies as well as the theoretical MoAs targeted by the strategy to better understand *how* and *why* implementation strategies impact implementation outcomes. Moreover, future research should continue to study and refine different combinations of BCTs that map onto a unique variety of MoAs, particularly for school-based contexts. For instance, feedback processes have been studied extensively as a potent MoA in health behavior and implementation of EBPs (Stormont et al., 2015), but exactly how these processes operate to promote change remains unknown. Moreover, the field of implementation science would benefit from increased specificity in describing the BCTs, which work to influence specific MoAs, and lead to improved implementation outcomes. This could be done by intentionally specifying, isolating, and varying the sequence of hypothesized BCTs that influence MoAs associated with implementation

strategies by using sequential multiple assignment randomized treatments (SMARTs) methodologies (Lei et al., 2012). Future research that uses SMART trials and other optimization designs will help develop more precise and effective ways of promoting teacher uptake and use of EBPs.

This study demonstrates that active implementation strategies (e.g., performance feedback, coaching, implementation planning) are effective to supporting teacher delivery of SEB EBPs with fidelity, but more research is needed testing strategies that seek to maintain levels of treatment integrity after implementation strategies are removed. Of the studies that collected follow-up data with no intentional maintenance strategy used, 63% demonstrated evidence of sustained adherence to the SEB EBP (determined through visual analysis of data, greater than 50% overlap to treatment phase). Moreover, only five studies incorporated specific strategies targeting teachers' maintenance of treatment integrity. These included systematic fading of supports after a set criterion was met (Digennaro et al., 2005; Gross et al., 2014, Hemmeter et al., 2011; Noell, Duhon, et al., 2002), and gradual fading from direct observation of treatment integrity to teacher self-monitoring (Oliver et al., 2015). Future school-based implementation research should consider developing and utilizing strategies specific to the sustainability of EBP implementation, such as dynamic fading of supports over time (Codding & Smyth, 2008; Digennaro et al., 2005). This broad concept can also be applied to a gradual release of consultative services from outside experts to school-based staff to enhance sustainability and systematic deployment of implementation strategies. Moreover, moving from consultation/coaching supports to self-monitoring of implementation can also be a feasible and effective way to sustain integrity as more costly and time intensive supports are withdrawn. This potentially has multiple other indirect effects including an internalizing of beliefs and values

related to treatment fidelity which leads, if multiplied across staff within a school building, to the propagation of a culture valuing and engaging in continual self-monitoring, reflection, and adjustment of practices.

A final consideration is the role that intervention adaptations have on the dynamic interplay of delivering EBPs with fidelity in real-world contexts. Adaptation is a key concept in implementation as it occurs for a variety of reasons, such as resource restrictions or fitting interventions to specific population demographics. Wiltsey-Stirman and colleagues (2019) developed the Framework for Reporting Adaptions and Modifications (FRAME), which is a framework for reporting adaptations to EBPs to better understand the process, types, and reasons that interventions are adapted. Utilizing FRAME during the implementation process can help end-users of EBPs make intentional decisions about fidelity-consistent adaptations to ensure that the effective practice components are maintained, whereas other components may be modified to better fit the context.

Implications for Implementation Practice

The findings from this study have implications for implementation practice broadly and within school settings. Similar to how schools organize their service delivery of academic and SEB supports for their students (e.g., MTSS), some school-based implementation researchers have advocated for a tiered approach to teacher implementation that considers the supports that all teachers need, supports that some teachers need, and supports that a few teachers need to successfully adopt and deliver EBPs (e.g., Myers et al., 2011; Sanetti et al., 2015). Although performance feedback is one of most studied and most effective school-based implementation strategies (e.g., Fallon et al., 2015), providing these supports can be costly and time consuming (Sanetti & Collier-Meek, 2015). Having a menu of tiered implementation strategies that can be

used to increase adherence would benefit schools; specifically, brief and feasible implementation strategies that target specific mechanisms of change should be developed to increase efficiency. We recommend that future researchers incorporate teachers into the development process so that strategies are built with the end-user in mind from the outset.

For practitioners, findings from this study indicate that one-off group-based didactic trainings are insufficient to facilitate teachers' adoption and sustained adherence to SEB EBPs. This is consistent with prior research that has demonstrated that didactic training alone results in low levels of implementation (Joyce & Showers, 2002). Moreover, among the studies included in this meta-analysis, only three used school-based staff as the actors of the implementation strategies. All others relied on external researchers who exert expert bases of power but not referent power (i.e., relationship and connection) that is more likely to be possessed by school-based consultants who are embedded within the school (Raven, 1993). However, Sanetti et al. (2013) found that internal consultants who provided evaluative performance-based feedback reported that their relationship with teachers changed, and that they were somewhat uncomfortable delivering performance feedback, as it was not something they had done in the past. Ultimately, there is a continued need for the development of pragmatic implementation strategies that school-based professionals who are embedded within schools can deliver effectively.

Although this study did not examine contextual factors that influence implementation, it is critical for implementation research to continue to explore factors associated with the organizational context of schools that influence teacher treatment integrity. Indeed, low implementation may be caused by the various contextual barriers that teachers face in schools (Collier-Meek et al., 2017). For example, school administrators play a vital leadership role in

implementation (Herlitz et al., 2020; Locke et al., 2019). They are able to allocate necessary resources (e.g., protected time, money, and materials) and hire skilled professionals who are able to participate on site-based teams and deliver consultative services that promote EBP implementation across all tiers of a multi-tiered framework (e.g., Sanetti & Kratochwill, 2009a). Because low implementation is associated with the contextual barriers teachers face in schools (e.g., lack of administrator support, negative school climate, policy), it is also important for future researchers to explore how barriers within other socio-ecological contexts impact teacher implementation of EBPs and how they interact with the individual implementers (Locke et al., 2019). The importance of focusing on multilevel influences on implementation success has been researched in other implementation contexts (e.g., Lewis et al., 2018) and has been highlighted as also being important for future research in school settings (Lyon & Bruns, 2019b). Therefore, it is recommended that agencies, such as schools or school districts, partner with implementation researchers or intermediaries who utilize implementation theory, models, frameworks, and strategies to not only develop and provide teacher professional development around the implementation of SEB supports to share knowledge and further develop the menu of effective and efficient implementation supports (Darling-Hammond et al., 2017; State et al., 2019), but also ensure that change is sustained by considering the various socioecological determinants of successful sustained implementation.

Conclusion

Student SEB needs are among teachers' most frequently reported classroom concerns. To ensure that classrooms function effectively, it is important that the best available practices with evidentiary support are routinely used to support student SEB outcomes in schools. However, the availability of EBPs alone does not ensure student SEB outcomes will improve. Achieving

beneficial outcomes ultimately depends on whether EBPs are consistently adopted and delivered with integrity. Implementation research in schools is a burgeoning field of inquiry, and it has the potential to improve both service delivery and student outcomes. What is known is that active implementation strategies are more effective than large-group, one-time trainings, and that multi-component strategies incorporating more BCTs that act on a wider variety of MoAs are more impactful than discrete, single-component strategies. Effective active implementation strategies include performance feedback, implementation planning, and prompts/reminders. The next phase of school-based implementation research should better specify the type of implementation strategies used by delineating the specific BCTs incorporated within the strategy to address hypothesized MoAs. Research that tailors strategies to need, while studying the temporal stage of the implementation process in which they are deployed as well as their sequence, should improve their efficacy and efficiency. It is the authors' hope that this meta-analysis will stimulate future school-based implementation research that draws from and contributes to the broader implementation science literature by further elucidating why, how, and for whom implementation strategies work.

References

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

Aarons, G.A., Ehrhart, M.G., & Farahnak, L.R., & Hurlburt, M. (2015). Leadership and organizational change for implementation (LOCI): A randomized mixed-method pilot study of a leadership and organization development intervention for evidence-based practice implementation. *Implementation Science, 10*(1), Article 11. https://doi.org/10.1186/s13012-014-0192-y

Aarons, G.A., Hurlburt, M., & Horwitz, S.M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administrative Policy in Mental Health, 38*(1), 4–23. https://doi.org/10.1007/s10488-010-0327-7

Adams, J., Hillier-Brown, F.C., Moore, H.J., Lake, A.A., Araujo-Soares, V., White, M., & Sumerbell, C. (2016). Searching and synthesising 'grey literature' and 'grey information' in public health: Critical reflections on three case studies. *Systematic Reviews,* 5, 164. https://doi.org/10.1186/s13643-016-0337-y

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

American Psychological Association Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist*, *61*(4), 271–285. https://doi.org/10.1037/0003-066X.61.4.271

Arumugam, A., Mehta P., & Baxter, G. (2020). Double-blind peer review of manuscripts: Opportunities, challenges, and way forward. *Physical Therapy Reviews*, *25*(1), 1–6, https://doi.org/10.1080/10833196.2019.1698161

Atkins, M.S., Rusch D., Mehta T., & Lakind, D. (2016). Future directions for dissemination and

implementation science: Aligning ecological theory and public health to close the research to practice gap. *Journal of Clinical Child and Adolescent Psychology, 45*(2), 215–226. https://doi.org/10.1080/15374416.2015.1050724

Barlow, D.H., & Nock, M.K. (2009). Why can't we be more idiographic in our research? *Perspectives on Psychological Science, 4*(1), 19–21. https://doi.org/10.1111/j.1745-6924.2009.01088.x

Beidas, R.S., & Kendall, P.C. (2010). Training therapists in evidence-based practice: A critical review of studies from a systems-contextual perspective. *Clinical Psychology, 17*(1), 1–30. https://doi.org/10.1111/j.1468-2850.2009.01187.x

Belur, J., Tompson, L., Thornton, A., & Simon, M. (2021). Interrater reliability in systematic review methodology: Exploring variation in coder decision-making. *Sociological Methods & Research*, *50*(2), 837–865. https://doi.org/10.1177/0049124118799372

Bergmann, S.C.J. (2018). *Establishing quality standards for applied behavior analytic skill-acquisition interventions: A translational model with undergraduate students*. [Unpublished doctoral dissertation]. University of Wisconsin–Milwaukee.

*Bethune, K.S. (2017). Effects of coaching on teachers' implementation of tier 1 school-wide positive behavioral interventions and support strategies. *Journal of Positive Behavior Interventions, 19*(3), 131–142. https://doi.org/10.1177%2F1098300716680095

Biggs, B.K., Venberg, E.M., Twemlow, S.W., Fonagy, P., & Dill, E.J. (2008). Teacher adherence and its relation to teacher attitudes and student outcomes in an elementary school-based violence prevention program. *School Psychology Review, 37*(4), 533–549. https://doi.org/10.1080/02796015.2008.12087866

Blewitt, C., Fuller-Tyszkiewicz, M., Nolan, A., Bergmeier, H., Vicary, D., Huang, T., McCabe,

P., McKay, T., & Skouteris, H. (2018). Social and emotional learning associated with universal curriculum-based interventions in early childhood education and care centers: A systematic review and meta-analysis. *JAMA Network Open*, *1*(8), e185727. https://doi.org/10.1186/1748-5908-7-37

Brackett, M.A., & Rivers, S.E. (2014). Transforming students' lives with social and emotional learning. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International Handbook of Emotions in Education* (pp. 368–388). Routledge/Taylor & Francis Group. https://www.routledge.com/International-Handbook-of-Emotions-in-Education/Pekrun-Linnenbrink-Garcia/p/book/9780415895026

Buckle, J. (n.d.). *The state of MTSS in education*. [Infographic]. https://www.panoramaed.com/blog/state-mtss-in-education-infographic

Byiers, B.J., Reichle, J., & Symons, F.J. (2012). Single-subject experimental design for evidence-based practice. *American Journal of Speech-Language Pathology*, *21*(4), 397–414. https://doi.org/10.1044/1058-0360(2012/11-0036)

Cane, J., O'Connor, D., & Michie, S. (2012). Validation of the theoretical domains framework for use in behavior change and implementation research. *Implementation Science*, *7*, Article 37. https://doi.org/10.1186/1748-5908-7-37

Carey, R.N., Connell Bohlen, L., Johnston, M., Rothman, A., de Bruin, M., Kelly, M.P., & Michie, S. (2018). Behavior change techniques and their MoAs: A synthesis of links described in published intervention literature. *Annals of Behavior Medicine, 53*(8), 693–707. https://doi.org/10.1093/abm/kay078

*Carter, D.R., & Van Norma, R.K. (2010). Class-wide positive behavior support in preschool: Improving teacher implementation through consultation. *Early Childhood Education*

Journal, 38, 279–288 https://doi.org/10.1007/s10643-010-0409-x

Catalano, R.F., Mazza, J.J., Herachi, T.W., Abbott, R.D., Hagerty, K.P., & Flemming, C.B. (2003). Raising healthy children through enhancing social development in elementary school: Results after 1.5 years. *Journal of School Psychology, 41*(2), 143–164. https://doi.org/10.1016/S0022-4405(03)00031-1

Chorpita, B.F., & Daleiden, E.L. (2009). Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Counseling and Clinical Psychology, 77*(5), 566–579. https://doi.org/10.1037/a0014565

Codding, R.S., & Smyth, C.A. (2008). Using performance feedback to decrease classroom transition time and examine collateral effects on academic engagement. *Journal of Educational Psychological Consultation, 18*(4), 325–345. https://doi.org/10.1080/10474410802463312

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. https://www.taylorfrancis.com/books/9780203771587

Collaborative for Academic, Social, and Emotional Learning. (2012). *2013 CASEL guide: Effective social and emotional learning programs—Preschool and elementary school edition*. Author. https://casel.org/wp-content/uploads/2016/01/2013-casel-guide-1.pdf

*Collier-Meek, M.A., Fallon, L.M., & Defouw, E.R. (2017). Toward feasible implementation support: E-mailed prompts to promote teachers' treatment integrity. *School Psychology Review 46*(4), 379–394. https://doi.org/10.17105/SPR-2017-0028.v46-4

Collier-Meek, M.A., Sanetti, L.M.H., & Boyle, A.M. (2016). Providing feasible implementation support: Direct training and implementation planning in consultation. *School Psychology*

Forum, *10*(1), 106–119. https://www.nasponline.org/publications/periodicals/spf/volume-10/volume-10-issue-1-(spring-2016)/providing-feasible-implementation-support-direct-training-and-implementation-planning-in-consultation

Connell Bohlen, L., Carey, R.N., de Bruin, M., Rothman, A., Johnston, M., Kelly, M.P., & Michie, S. (2018). Links between behaviour change techniques and MoAs: An expert consensus study. *Annals of Behavioral Medicine, 53*(8). 708–720. https://doi.org/10.31234/osf.io/fge86

Cook, B.G., Smith, G.J., & Tankersley, M. (2012). *Evidence-based practices in education.* In K.R. Harris, S. Graham, T. Urdan, C.B. McCormick, G.M. Sinatra, & J. Sweller (Eds.), *APA handbooks in psychology®. APA educational psychology handbook: Vol. 1. Theories, constructs, and critical issues* (p. 495–527). American Psychological Association. https://doi.org/10.1037/13273-017

Cook, C.R., Burns, M., Browning-Wright, D., & Gresham, F.M. (2010). *Transforming school psychology in the RTI era: A guide for administrators and school psychologists*. LRP Publications. https://www.shoplrp.com/product_p/300617.htm

Cook, C.R., Dart, E., Collins, T., Restori, A.,Vance, M., & Fitts, P. (2013). Co-occurring reading and behavior problems: Transactional relationship or not? Implications for intervention. *Behavioral Disorders, 38*(1), 38–56. https://www.jstor.org/stable/43153567?seq=1

Cook, C.R., Fiat, A., Larson, M., Daikos, C., Slemrod, T., Holland, E.A., ... Renshaw, T. (2018). Positive greetings at the door: Evaluation of a low-cost, high-yield proactive classroom management strategy. *Journal of Positive Behavior Interventions*, *20*(3), 149–159. https://doi.org/10.1177%2F1098300717753831

Cook, C.R., Lyon, A.R., Locke, J., Waltz, T.J., & Powell, B.J. (2019). Adapting a compilation of

implementation strategies to advance school-based implementation research and practice. *Prevention Science, 20*(6), 914–935. https://doi.org/10.1007/s11121-019-01017-1

Crawley, S.A., Podell, J.L., Beidas, R.S., Braswell, L., & Kendall, P.C. (2010). Cognitive-behavioral therapy. In K. Dobson (Ed.), *Handbook of cognitive-behavioral therapies* (pp. 375–410). Guilford Press. https://www.guilford.com/books/Handbook-of-Cognitive-Behavioral-Therapies/Dobson-Dozois/9781462538584

Damschroder, L.J., Aron, D.C., Keith, R.E., Kirsh, S.R., Alexander, J.A., & Lowery, J.C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science, 4,* Article 50. https://doi.org/10.1186/1748-5908-4-50

Darling-Hammond, L., Hyler, M.E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute. https://learningpolicyinstitute.org

*Dart, E.H., Cook, C.R., Collins, T.A., Gresham, F.M., & Chenier, J.S. (2012). Test driving interventions to increase treatment integrity and student outcomes. *School Psychology Review, 41*(4), 467–481. https://doi.org/10.1080/02796015.2012.12087500

*Digennaro, F.D., Martens, B.K., & McIntyre, L.L. (2005). Increasing treatment integrity through negative reinforcement: Effects of teacher and student behavior. *School Psychology Review, 34*(2), 220–231. https://doi.org/10.1080/02796015.2005.12086284

Dombrowski, S.U., Knittle, K., Avenell, A., Araújo-Soares, V., & Sniehotta, F.F. (2014). Long term maintenance of weight loss with non-surgical interventions in obese adults: Systematic review and meta-analyses of randomised controlled trials. *British Medical Journal, 348,* Article g2646. http://www.bmj.com/content/348/bmj.g2646.

Domitrovich, C.E., Durlak, J.A., Staley, K.C. & Weissberg, R.P. (2017). Social- emotional

competence: An essential factor for promoting positive adjustment and reducing risk in school children. *Child Development, 88*(2), 408–416. https://doi.org/10.1111/cdev.12739

Drevon, D., Fursa, S.R., & Malcolm, A.L. (2017). Intercoder reliability and validity of WebPlotDigitizer in extracting graphed data. *Behavior Modification, 41*(2), 323–339. https://doi.org/10.1177/0145445516673998

*Dufrene, B.A., Parky, K., Menousek, K., Zhou, Q., Harpole, L.L., & Olmi, D.J. (2012). Direct behavioral consultation in Head Start to increase teacher use of praise and effective instruction delivery. *Journal of Educational and Psychological Consultation, 22*(3),159–186. https://doi.org/10.1080/10474412.2011.620817

Durlak J.A., & DuPre, E.P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*(3–4), 327–350. https://doi.org/10.1007/s10464-008-9165-0

Durlak, J.A., Weissberg, R.P., Dymnicki, A.B., Taylor, R.D., & Schellinger, K.B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

Egger, M., Smith, G.D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*, 629. https://doi.org/10.1136/bmj.315.7109.629

Every Student Succeeds Act of 2015, Pub. L. No. 114-95 § 114 Stat. 1177. https://www.congress.gov/bill/114th-congress/senate-bill/1177/text

Fallon, L.M., Collier-Meek, M.A., Maggin, D.M., Sanetti, L.M.H., & Johnson, A.H (2015). Is performance feedback an evidence-based practice? A systematic review and evaluation

based on single case research. *Exceptional Children, 81*(2), 227–246. https://doi.org/10.1177/0014402914551738

Fernandez, M.E., ten Hoor, G.A., van Lieshout, S., Rodriguez, S.A., Beidas, R.S., Parcel, G., & Kok, G. (2019). Implementation mapping: Using intervention mapping to develop implementation strategies. *Frontiers in Public Health, 7* Article: 158. https://doi.org/10.3389/fpubh.2019.00158

Fisher, Z., & Tipton, Z. (2015). robumeta: An R package for robust variance estimation in meta-analysis. *Mathematics*. https://arxiv.org/pdf/1503.02220.pdf

Fiske, K. (2008). Treatment integrity of school-based behavior analytic interventions: A review of the research. *Behavior Analysis in Practice, 1*(2), 19–25. https://dx.doi.org/10.1007%2FBF03391724

Fjeldsoe, B., Neuhaus, M., Winkler, E., & Eakin, E. (2011). Systematic review of maintenance of behavior change following physical activity and dietary interventions. *Health Psychology, 30*(1), 99–109. https://doi.org/10.1037/a0021974

*Fullerton, E.K., Conroy, M.A., & Correa, V.I. (2009). Early childhood teachers' use of specific praise statements with young children at risk for behavioral disorder. *Behavioral Disorders, 34*(3), 118–135. https://doi.org/10.1177%2F019874290903400302

George H.P., Kincaid, D. & Pollard-Sage J. (2009). Primary-tier interventions and supports. In W. Sailor, G. Dunlap, G. Sugai, & R. Horner (Eds.), *Handbook of positive behavior support: Issues in clinical child psychology* (pp. 375-394). Springer. https://www.springer.com/gp/book/9780387096315

Greenberg, M.T., Weissberg, R.P., O'Brien, M.U., Zins, J.E., Fredericks, L., Resnik, H., & Elias, M.J. (2003). Enhancing school-based prevention and youth development through

coordinated social, emotional, and academic learning. *American Psychologist, 58*(6–7), 466–474. https://doi.org/10.1037/0003-066X.58.6-7.466

*Gross, T.J., Duhon, G.J., & Doerksen-Klopp, B. (2014). Enhancing treatment integrity maintenance through fading with indiscriminable contingencies. *Journal of Behavioral Education, 23,* 108–131. https://doi.org/10.1007/s10864-013-9185-1

Hawken, L.S., Adolphson, S.L., Macleod, K.S., & Schumann, J. (2009). Secondary-tier interventions and supports. In W. Sailor, G. Dunlap, G. Sugai, & R. Horner (Eds.), *Handbook of positive behavior support: Issues in clinical child psychology* (pp. 395-420). Springer. https://link.springer.com/chapter/10.1007/978-0-387-09632-2_17#citeas

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

Hedges, L.V., Tipton, E., & Johnson, M.C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*(1), 39–65. https://doi.org/10.1002/jrsm.5

*Hemmeter, M.L., Snyder, P., Kinder, K., & Artman, K. (2011). Impact of performance feedback delivered via electronic mail on preschool teachers' use of descriptive praise. *Early Childhood Research Quarterly, 26*(1), 196–209. https://doi.org/10.1016/j.ecresq.2010.05.004

Herlitz, L., MacIntyre, H., Osborn, T. & Bonell, C. (2020). The sustainability of public health interventions in schools: A systematic review. *Implementation Science, 15,* Article 4. https://doi.org/10.1186/s13012-019-0961-8

*Hundert, J., & Hopkins, B. (1992). Training supervisors in a collaborative team approach to promote peer interaction of children with disabilities in integrated preschools. *Journal of Applied Behavior Analysis, 25*(2), 385–400. https://dx.doi.org/10.1901%2Fjaba.1992.25-385

Individuals with Disabilities Education Act, 20 U.S.C. § 1400 (2004). https://sites.ed.gov/idea/

Johnston, M., Carey, R.N., Connell Bohlen, L., Johnston, D.W., Rothman, A., de Bruin, M., … Michie, S. (2018). Linking behavior change techniques and MoAs: Triangulation of findings from literature synthesis and expert consensus. *Translational Behavioral Medicine, 11*(5), 1049–1065. https://doi.org/10.31234/osf.io/ur6kz

Joyce, B.R., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Association for Supervision & Curriculum Development (ASCD).

Kessler, R., & Glasgow, R.E. (2011). A proposal to speed translation of healthcare research into practice. *American Journal of Preventive Medicine, 40(6)*, 637–644. https://doi.org/10.1016/j.amepre.2011.02.023

*Kleinert, W.L., Silva, M.R., Codding, R.S., Feinberg, A.B., & St. James, P.S. (2017). Enhancing classroom management using the classroom check-up consultation model with in-vivo coaching and goal setting components. *School Psychology Forum, 11*(1), 5–19. https://journals.sagepub.com/doi/pdf/10.1177/0198742917704648

Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research, 86*(3), 643–680. https://doi.org/10.3102/0034654315626799

Kowalski, T., & Lasley, T. (2009). *Handbook of data-based decision making in education.*

Routledge. https://doi.org/10.3102/0034654315626799

Krasny-Pacini, A., & Evans, J. (2018). Single-case experimental designs to assess intervention effectiveness in rehabilitation: A practical guide. *Annals of physical and rehabilitation medicine*, *61*(3), 164–179. https://doi.org/10.1016/j.rehab.2017.12.002

Kratochwill, T.R., Hitchcock, J., Horner, R.H., Levin, J.R., Odom, S.L., Rindskopf, D.M., & Shadish, W.R. (2010). *Single-case designs technical documentation*. http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Kulinskaya, E., & Dollinger, M.B. (2015). An accurate test for homogeneity of odds ratios based on Cochran's *Q*-statistic. *BMC Med Res Methodol, 15,* 49. https://doi.org/10.1186/s12874-015-0034-x

Krause, J., Van Lieshout, J., Klomp, R., Huntink, E., Aakhus, E., Flottorp, S., Jaeger, C., Steinhaeuser, J., Godycki-Cwirko, M., Kowalczyk, A., Agarwal, S., Wensing, W., & Baker, R. (2014). Identifying determinants of care for tailoring implementation in chronic diseases: An evaluation of different methods. *Implementation Science, 9,* Article 102. https://doi.org/10.1186/s13012-014-0102-3

Lane, J.D., & Gast, D.L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation*, *24*(3–4), 445–463. https://doi.org/10.1080/09602011.2013.815636

Larson, M.F., Cook, C.R., Brewer, S.K., Pullmann, M.D., Hamlin, C., Merle, J.L., Duong, M., Gaias, L., Sullivan, M., Morrell, N., Kulkarni, T., Weeks, M., & Lyon, A.R. (2020). Examining the effects of a brief, group-based motivational implementation strategy on mechanisms of teacher behavior change. *Prevention Science*, *22*(6), 722–736. https://doi.org/10.1007/s11121-020-01191-7

Larson, M., Cook, C.R., Fiat, A., & Lyon, A.R. (2018). Stressed teachers don't make good implementers: Examining the interplay between stress reduction and intervention fidelity. *School Mental Health, 10,* 61–76. https://doi.org/10.1007/s12310-018-9250-y

Ledford, J.R., & Gast, D.L. (Eds.). (2018). *Single case research methodology: Applications in special education and behavioral sciences*. Routledge. https://doi.org/10.4324/9781315150666

Lei, H., Nahum-Shani, I., Lynch, K., Oslin, D., & Murphy, S. (2012): A "SMART" design for building individualized treatment sequences. *Annual Review of Clinical Psychology, 8,* 21–48. https://dx.doi.org/10.1146%2Fannurev-clinpsy-032511-143152

Lewis, C.C., Klasnja, P., Powell, B.J., Lyon, A.R., Tuzzio, L., Jones, S., . . . Weiner, B. (2018). From classification to causality: Advancing understanding of mechanisms of change in implementation science. *Frontiers in Public Health, 6,* Article 136. https://dx.doi.org/10.3389%2Ffpubh.2018.00136

Liddy, C., Wiens, M., & Hogg, W. (2011). Methods to achieve high interrater reliability in data collection from primary care medical records. *Annals of Family Medicine*, *9*(1), 57–62. https://doi.org/10.1370/afm.1195

Lipsey, M.W., & Wilson, D.B. (2001). *Practical meta-analysis*. Sage.

Lobo, M.A., Moeyaert, M., Baraldi Cunha, A., & Babik, I. (2017). SCED, analysis, and quality assessment for intervention research. *Journal of Neurologic Physical Therapy*, *41*(3), 187–197. https://doi.org/10.1097/NPT.0000000000000187

Lochman, J.E., & Wells, K.C. (2004). The Coping Power Program for preadolescent aggressive boys and their parents: Outcome effects at the 1-year follow-up. *Journal of Consulting and Clinical Psychology, 72*(4), 571–578. http://dx.doi.org/10.1037/0022-006X.72.4.571

Locke, J., Lawson, G.M., Beidas, R.S., Aarons, G.A., Xie, M., Lyon, A.R., Stahmer, A., . . . Mandell, D.S. (2019). Individual and organizational factors that affect implementation of evidence-based practices for children with autism in public schools: A cross-sectional observational study. *Implementation Science, 14*, Article 29. https://dx.doi.org/10.1186%2Fs13012-019-0877-3

Long, A.C., Hagermoser Sanetti, L.M., Collier-Meek, M.A., Gallucci, J., Altschaefl, M., & Kratochwill, T.R. (2016). An exploratory investigation of teachers' intervention planning and perceived implementation barriers. *Journal of School Psychology*, *55*, 1–26. https://doi.org/10.1016/j.jsp.2015.12.002

Lyon, A.R., Cook, C.R., Duong, M.T., Nicodimos, S., Pullmann, M.D., Brewer, S.K., Gaias, L.M., & Cox, S. (2019). The influence of a blended, theoretically-informed pre-implementation strategy on school-based clinician implementation an evidence-based trauma intervention. *Implementation Science,14*(1), Article 54. https://doi.org/10.1186/s13012-019-0905-3

Lyon, A.R., & Bruns, E.J. (2019a). User-centered redesign of evidence-based psychosocial interventions to enhance implementation—Hospitable soil or better seeds? *JAMA Psychiatry, 76*(1), 3–4. https://doi.org/10.1001/jamapsychiatry.2018.3060

Lyon, A.R., & Bruns, E.J. (2019b). From evidence to impact: Joining our best school mental health practices with our best implementation strategies. *School Mental Health, 11*, 106–114. https://doi.org/10.1007/s12310-018-09306-w

Lyon, A.R., Pullmann, M.D., Walker, S.C., & D'Angelo, G. (2017). Community-sourced intervention programs: Review of submissions in response to a statewide call for "promising practices". *ADM Policy Ment Health, 44*, 16-28.

https://link.springer.com/content/pdf/10.1007/s10488-015-0650-0.pdf

Makel, M.C., & Plucker, J.A. (2014). Facts are more important that novelty: Replication in the education sciences. *Educational Researcher, 43*(6), 304–316. https://doi.org/10.3102%2F0013189X14545513

Mathur, S.R., Kavale, K.A., Quinn, M.M., Forness, S.R., & Rutherford, R.B. (1998). Social skills interventions with students with emotional and behavioral problems: A quantitative synthesis of single-subject research. *Behavioral Disorders*, *23*(3), 193–201. https://doi.org/10.1177/019874299802300306

McClain, M.B., Callan, G.L., Harris, B., Floyd, R.G., Haverkamp, C.R., Golson, M.E., Longhurst, D.N., & Benallie, K.J. (2021). Methods for addressing publication bias in school psychology journals: A descriptive review of meta-analyses from 1980 to 2019. *Journal of School Psychology*, *84*, 74–94.

McKenna, J.W., Flower, A., & Ciullo, S. (2014). Measuring fidelity to improve intervention effectiveness. *Intervention in School and Clinic, 50*(1), 15–21. https://doi.org/10.1177/1053451214532348

*McKenney, E.W., Waldron, N., & Conroy, M. (2013). The effects of training and performance feedback during behavioral consultation on general education middle school teachers' integrity to functional analysis procedures. *Journal of Educational and Psychological Consultation, 23*(1), 63–85. https://doi.org/10.1080/10474412.2013.757152

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

Michie, S., Carey, R., Johnston, M., Rothman, A.J., Kelly, M., Davidson, K., & de Bruin, M. (2016). From theory-inspired to theory-based interventions: Linking behaviour change

techniques to their MoAs. *European Journal of Health Psychology*, *18*, Article 395. https://doi.org/10.1007/s12160-016-9816-6

Michie, S., Carey, R., Johnston, M., Rothman, A., Bruin, M., Kelly, M. & Connell, L. (2018). From theory-inspired to theory-based interventions: A protocol for developing and testing a methodology for linking behavior change techniques to theoretical MoAs. *Annals of Behavioral Medicine, 52*(6), 501–512. https://doi.org/10.1007/s12160-016-9816-6

Michie, S., Fixsen, D., Grimshaw, J.M., & Eccles, M.P. (2009). Specifying and reporting complex behaviour change interventions: The need for a scientific method. *Implementation Science, 4*(1), Article 40. https://doi.org/10.1186/1748-5908-4-40

Michie, S., Johnston, M., Rothman, A.J., de Bruin, M., Kelly, M.P., Carey, R.N., . . . Zink, S. (2021). Developing an evidence-based online method of linking behaviour change techniques and theoretical mechanisms of action: A multiple methods study. *National Institute for Health Research, 9*(1). https://www.ncbi.nlm.nih.gov/books/NBK567047/pdf/Bookshelf_NBK567047.pdf

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J, Hardeman, W., . . . Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine, 46*(1), 81–95. https://doi.org/10.1007/s12160-013-9486-6

Moore, E., Gaskin, C., & Indig, D. (2013). Childhood maltreatment and post-traumatic stress disorder among incarcerated young offenders. *Child Abuse and Neglect, 37*(10), 861–870. https://doi.org/10.1016/j.chiabu.2013.07.012

Moullin, J.C., Dickson, K.S., Stadnick, N.A., Rabin, B., & Aarons, G.G. (2019). Systematic

review of the exploration, preparation, implementation, sustainment (EPIS) framework. *Implementation Science, 14,* Article 1. http://doi.org/10.1186/s13012-018-0842-6

*Myers, D.M., Simonsen, B., & Sugai, G. (2011). Increasing teachers' use of praise with a response-to-intervention approach. *Education and Treatment of Children, 34*(1), 35–59. https://doi.org/10.1353/etc.2011.0004

Nadeem, E., Olin, S.S., Hill, L.C., Hoagwood, K.E., & Horwitz, S.M. (2013). Understanding the components of quality improvement collaboratives: A systematic literature review. *Milbank Q, 91*(2), 354–94. https://doi.org/10.1111/milq.12016

National Scientific Council on the Developing Child (2004). *Children's emotional development is built into the architecture of their brains*. Center on the Developing Child, Harvard University. https://developingchild.harvard.edu/science/national-scientific-council-on-the-developing-child/

Noell, G.H. (2008). Research examining the relationships among consultation process, treatment integrity, and outcomes. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation: Empirical foundations for the field* (pp. 315–334). Erlbaum. https://doi.org/10.4324/9780203133170

*Noell, G.H., Duhon, G.J., Gatti, S.L., & Connell, J.E (2002). Consultation, follow-up, and implementation of behavior management interventions in general education. *School Psychology Review, 31*(2), 217–234. https://doi.org/10.1080/02796015.2002.12086152

Noell, G.H., Gansle, K.A., Meyeres, J.L., Knox, R.M., Mintz, J.C., & Dahir, A. (2014) Improving treatment plan implementation in schools: A meta-analysis of single subject design studies. *Journal of Behavioral Education, 23*(1), 168–191. https://doi.org/10.1007/s10864-013-9177-1

Noell, G.H., Gresham, F.M., & Gansle, K.A. (2002). Does treatment integrity matter? A preliminary investigation of instructional implementation and mathematics performance. *Journal of Behavioral Education, 11*(1), 51–67. https://doi.org/10.1023/A:1014385321849

Noell, G.H., Witt, J.C., Gilbertson, D.N., Ranier, D.D., & Freeland, J.T. (1997). Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *School Psychology Quarterly, 12*(1), 77–88. https://doi.org/10.1037/h0088949

Noell, G.H., Witt, J.C., Slider, N.J., Connell, J.E., Gatti, S.L., Williams, K.J., Koenig, J. L., Resetar, J.L., & Duhon, G.J. (2005) Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review*, *34*(1), 87–106, DOI: 10.1080/02796015.2005.1208627

O'Donnell, C.L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, *78*(1), 33–84. https://doi.org/10.3102/0034654307313793

*Oliver, R.M., Wehby, J.H., & Nelson, J.R. (2015). Helping teachers maintain classroom management practices using a self-monitoring checklist. *Teaching and Teacher Education, 51,* 113–120. https://doi.org/10.1016/j.tate.2015.06.007

Orwin, R.G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics, 8(2),* 157–159. https://doi.org/10.2307/1164923

Owens, J.S., Lyon, A.R., Brandt, N.E., Warner, C.M., Nadeem, E., Spiel, C., & Wagner, M. (2014). Implementation science in school mental health: Key constructs in a developing research agenda. *School Mental Health,* 6(2), 99–111. https://doi.org/10.1007/s12310-

013-9115-3

Parker, R.I., & Vannest, K. (2009). An improved effect size for single-case research: nonoverlap of all pairs. *Behavior therapy, 40*(4), 357–367. https://doi.org/10.1016/j.beth.2008.10.006

Parker, R.I., Vannest, K.J., Davis. J.L., & Sauber, S.B (2011). Combining nonoverlap and trend for single case research: Tau-U. *Behavior Therapy*, *42*(2), 284–299. https://doi.org/10.1016/j.beth.2010.08.006

Petersilia, J. (1990). Conditions that permit intensive supervision to survive. *Crime & Delinquency*, *36*(1), 126–145. https://doi.org/10.1177/0011128790036001009

Powell, J.P., McMillen, J.C., Proctor, E.K., Carpenter, C.R., Griffey, R.T., Bunger, A.C., Glass, J.E., & York., J.L. (2012). A compilation of strategies for implementing clinical innovations in health and mental health. *Medical Care Research and Review, 62*(2), 123–157. https://doi.org/10.1177/1077558711430690

Powell, B.J., Fernandez, M.E., Williams, N.J., Aarons, G.A., Beidas, R.S., Lewis, C.C., ... & Weiner, B.J. (2019). Enhancing the impact of implementation strategies in healthcare: a research agenda. *Frontiers in Public Health, 7,* Article 3. https://doi.org/10.3389/fpubh.2019.00003

Powell, B.J., Proctor, E.K., & Glass, J.E. (2014). A systematic review of strategies for implementing empirically supported mental health interventions. *Research on Social Work Practice, 24*(2), 192–212. https://doi.org/10.1177/1049731513505778

Powell B.J., Waltz T.J., Chinman M.J., Damschroder, L.J., Smith, J.L., Mattieu, M.M., … Kirchner, J.E. (2015). A refined compilation of implementation strategies: Results from the Expert Recommendations for Implementing Change (ERIC) project. *Implementation Science, 10*(21), 1–14. https://doi.org/10.1186/s13012-015-0209-1

Proctor, E.K., Powell, B.J., & McMillen, J.C. (2013). Implementation strategies: Recommendations for specifying and reporting. *Implementation Science, 8,* Article 139. https://doi.org/10.1186/1748-5908-8-139

Proctor, E.K., Silmere, H., Maghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(2), 65–76. https://doi.org/10.1007/s10488-010-0319-7

Pustejovsky, J.E. (2016, March 23). Tau-U. *https://www.jepusto.com/tau-u/*

Putnam, R., McCart, A., Griggs, P., & Choi, J.H. (2009). Implementation of schoolwide positive behavior supports in urban settings. In W. Sailor, G. Dunlap, G. Sugai, & R. Horner (Eds.), *Handbook of positive behavior support: Issues in Clinical child Psychology* (pp. 443-463). Springer. https://doi.org/10.1007/978-0-387-09632-2_19

Ramos-Vidal, I., Palacio, J., Villamil, I. & Uribe, A. (2020). Examining the effect of the individual characteristics of implementers and the interaction of multiple relationships on the structure of psychosocial intervention teams. *Implementation Science, 15***,** Article 69. https://doi.org/10.1186/s13012-020-01032-9

Ransford, C.R., Greenberg, M.T., Domitrovich, C.E., Small, M., & Jacobson, L. (2009). The role of teachers' psychological experiences and perceptions of curriculum supports on the implementation of a social and emotional learning curriculum. *School Psychology Review, 38*(4), 510–532. https://psycnet.apa.org/record/2010-00081-005

Raven, B.H. (1993). The bases of power: Origins and recent developments. *Journal of Social Issues, 49*(4), 227–251. https://doi.org/10.1111/j.1540-4560.1993.tb01191.x

Reichow, B., Volkmar, F.R., & Cicchetti, D.V. (2008). Development of the evaluative method for evaluating and determining evidence- based practices in autism. *Journal of Autism and Developmental Disorders, 38*(7), 1311–1319. https://doi.org/10.1007/s10803-007-0517-7

*Reinke, W.M., Palmer, T.L., & Martin, E. (2007). The effect of visual performance feedback on teacher use of behavior-specific praise. *Behavior Modification, 31*(3), 247–263. https://doi.org/10.1177/0145445506288967

*Reinke, W.M., Palmer, T.L., & Merrell, K. (2008). The Classroom Check-Up: A classwide teacher consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review, 37*(3), 315–332. https://doi.org/10.1080/02796015.2008.12087879

Reinke, W.M., Stormont, M., Herman, K.C., Puri, R., & Goel, N. (2011). Supporting children's mental health in schools: Teacher perceptions of needs, roles, and barriers. *School Psychology Quarterly, 26*(1), 1–13. https://doi.org/10.1037/a0022714

*Riley-Tillman, T.C., & Eckert, T.L. (2001). Generalization programming and school-based consultation; An examination of consultees' generalization of consultation-related skills. *Journal of Educational and Psychological Consultation, 12*(2), 217–241. https://doi.org/10.1207/S1532768XJEPC1203_03

*Rispoli, M., Ninci, J., Burke, M.D., Zaini, S., Hatton, H., & Sanchez, L. (2015). Evaluating the accuracy of results for teacher implemented trial-based functional analyses. *Behavior Modification, 39*(5), 627–653. https://doi.org/10.1177/0145445515590456

*Rodriguez, B.J., Sheldon, M.S., Loman, M.A., & Horner, R.H. (2009). A preliminary analysis of effects of coaching feedback on teacher implementation fidelity of First Step to

Success. *Behavior Analysis in Practice, 2*(2), 11–21. https://doi.org/10.1007/BF03391744

Rohatgi, A. (2019). *WebPlotDigitizer* (Version 4.2). Retrieved from https://automeris.io/WebPlotDigitizer

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rothstein, H.R., Sutton, A.J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Wiley. https://www.meta-analysis.com/downloads/PBPreface.pdf

*Sanetti, L.M.H., & Collier-Meek, M.A. (2015) Data-driven delivery of implementation supports in a multi-tiered framework: A pilot study. *Psychology in the Schools, 52*(8), 815–828. https://doi.org/10.1002/pits.21861

*Sanetti, L.M.H., & Collier-Meek, M.A. (2013). Increasing the rigor of procedural fidelity assessment: An empirical comparison of direct observation and permanent product review methods. *Journal of Behavioral Education, 23*, 60–88. https://doi.org/10.1007/s10864-013-9179-z

*Sanetti, L.M.H., Collier-Meek, M.A., Long, A.C.J., Byron, J., & Kratochwill, T.R. (2015). Increasing teacher treatment integrity of behavior support plans through consultation and implementation planning. *Journal of School Psychology, 53*(3), 209–229. https://doi.org/10.1016/j.jsp.2015.03.002

*Sanetti, L.M.H., Collier-Meek, M.A., Long, A.C.J., Kim, J.S., & Kratochwill, T.R. (2014). Using implementation planning to increase teachers' adherence and quality to behavior support plans. *Psychology in the Schools, 51*(8), 879–895. https://doi.org/10.1002/pits.21787

Sanetti, L.M.H., Fallon, L.M., & Collier- Meek, M.A. (2013). Increasing teacher treatment integrity through performance feedback provided by school personnel. *Psychology in the Schools, 50*(2) 134–150. https://doi.org/10.1177/0014402914551738

*Sanetti, L.M.H., & Kratochwill, T.R. (2011). An evaluation of the treatment integrity planning protocol and two schedules of integrity self-report: Impact on implementation and report accuracy. *Journal of Educational and Psychological Consultation, 21*(4), 284–308. https://doi.org/10.1080/10474412.2011.620927

Sanetti, L.M.H., & Kratochwill, T.R. (2009a). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38*(4), 445–459.

Sanetti, L.M.H., & Kratochwill, T.R. (2009b). Treatment integrity assessment in the schools: An evaluation of the treatment integrity planning Protocol. *School Psychology Quarterly, 24*(1), 24–35. https://doi.org/10.1037/a0015431

Sanetti, L.M.H., & Kratochwill, T.R. (2007). Treatment integrity in behavioral consultation: Measurement, promotion, and outcomes. *International Journal of Behavioral Consultation and Therapy, 4*(1), 95–114. http://dx.doi.org/10.1037/h0100835

Schmucker, C.M., Blümle, A., Schell, L.K., Schwarzer, G., Oeller, P., Cabrera, L., von Elm, E., Briel, M., Meerpohl, J.J., & OPEN Consortium (2017). Systematic review finds that study data not published in full text articles have unclear impact on meta-analyses results in medical research. *PloS one, 12*(4), Article e0176210. https://doi.org/10.1371/journal.pone.0176210

Schulte, A.C., Easton, J.E., & Parker, J. (2009). Advances in treatment integrity research: Multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treatment integrity. *School Psychology Review, 38*(4), 460–475.

Schwarzer, R. (2008). Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. *Applied Psychology: An International Review, 57*(1), 1–29. https://doi.org/10.1111/j.1464-0597.2007.00325.x

Shadish, W.R., Hedges, L.V., & Pustejovsky, J.E. (2014). Analysis and meta-analysis of SCEDs with a standardized mean difference statistic: A primer and applications. *Journal of School Psychology, 52*(2), 123–147. https://doi.org/10.1016/j.jsp.2013.11.005

Simmons, A. (2013). Suspensions and expulsions contribute to school dropouts. *Journal of Educational Controversy, 7*(1), Article 14. https://cedar.wwu.edu/jec/vol7/iss1/14

*Simonsen, B., Freeman, J., Dooley, K., Maddock, E., Kern, L., & Myers, D. (2017). Effects of targeted professional development of teachers' specific praise rates. *Journal of Positive Behavior Interventions, 19*(1), 37–47. https://doi.org/10.1177/1098300716637192

Skiba, R.J., & Rausch, M.K. (2006). Zero tolerance, suspension, and expulsion: Questions of equity and effectiveness. In C.M. Evertson & C.S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 1063–1089). Lawrence Erlbaum Associates Publishers.

Smith, J.D. (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods, 17*(4), 510–550. https://doi.org/10.1037/a0029312

*Smith, S.C., Lewis, T.J., & Stormont, M. (2011). The effectiveness of two universal behavioral supports for children with externalizing behavior in Head Start classroom. *Journal of Positive Behavior Interventions, 13*(3), 133–143. https://doi.org/10.1177/1098300710379053

Snyder, T.D., & Dillow, S.A. (2015). *Digest of education statistics 2013* (NCES 2015-011). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. https://nces.ed.gov/pubs2015/2015011.pdf

Solomon, B.G., Klein, S.A., & Politylo, B.C. (2012). The effect of performance feedback on teachers' treatment integrity: A meta-analysis of the single case literature. *School Psychology Review, 41*(2), 160–175. https://doi.org/10.1080/02796015.2012.12087518

Squires, J.E., Sullivan, K., Eccles, M.P., Worswick, J., & Grimshaw, J.M. (2014). Are multifaceted interventions more effective than single component interventions in changing healthcare professionals' behaviours? An overview of systematic reviews. *Implementation Science, 9,* Article 152. https://doi.org/10.1186/s13012-014-0152-6

State, T.M., Simonsen, B., Hirn, R.G., & Wills, H. (2019). Bridging the research-to-practice gap through effective professional development for teachers working with students with emotional and behavioral disorders. *Behavioral Disorders, 44*(2), 107–116. https://doi.org/10.1177%2F0198742918816447

Sterne, J.A.C., Sutton, A.J., Ioannidis, J.P.A., Terrin, N., Jones, D.R., Lau, J., . . . Higgins, J.P.T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ, 343,* Article d4002. https://doi.org/10.1136/bmj.d4002

Stormont, M., Reinke, W.M., Newcomer, L., Marchese, D., & Lewis, C. (2015). Coaching teachers' use of social behavior interventions to improve children's outcomes: A review of the literature. *Journal of Positive Behavior Interventions*, *17*(2), 69–82. https://doi.org/10.1177/1098300714550657

*Stormont, M.,A., Smith, S.A., & Lewis, T.J. (2007). Teacher implementation of precorrection and praise statements in Head Start classrooms as a component of a program-wide system of positive behavior support. *Journal of Behavioral Education, 16*, 280–290. https://doi.org/10.1007/s10864-007-9040-3

Tabak, R.G., Khoong, E.C., Chambers, D.A., & Brownson, R.C. (2012). Bridging research and practice: Models for dissemination and implementation research. *American Journal of Preventive Medicine, 43*(3), 337–350. https://doi.org/10.1016/j.amepre.2012.05.024

Taylor, R.D., Oberle, E., Durlak, J.A., & Weissberg, R.P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development, 88*(4), 1156–1171. https://doi.org/10.1111/cdev.12864

Townsend, L., Flisher, A.J., & King, G. (2007). A systematic review of the relationship between high school dropout and substance use. *Clinical Child and Family Psychology Review, 10*(4), 295–317. https://doi.org/10.1007/s10567-007-0023-7

Turri, M., Mercer, S., McIntosh, K., Nese, R., Strickland-Cohen, M., Hoselton, R. (2016). Examining barriers to sustained implementation of school-wide prevention practices. *Assessment for Effective Intervention, 42*(1), 6–17. https://doi.org/10.1177/1534508416634624

van Buuren, S. (2018). *Flexible imputation of missing data.* Chapman & Hall. https://stefvanbuuren.name/fimd/

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. http://dx.doi.org/10.18637/jss.v045.i03

Vannest, K.J., & Ninci, J. (2015). Evaluating intervention effects in single case research designs. *Journal of Counseling and Development, 93*(4), 403–411. https://doi.org/10.1002/jcad.12038

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

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. http://dx.doi.org/10.18637/jss.v036.i03

Waldfogel, J., Garfinkel, I., & Kelly, B. (2007). Welfare and the costs of public assistance. In C.R. Belfield & H.M. Levin (Eds.), *The price we pay: Economic and social consequences of inadequate education* (pp. 160–176). Brookings Institution. https://www.brookings.edu/book/the-price-we-pay/

Webster, J. (2019, February 16). *What is general education? ThoughtCo*. https://www.thoughtco.com/general-education-glossary-term-3110863

Westen, D., Novotny, C.M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin, 130*(4), 631–663. https://doi.org/10.1037/0033-2909.130.4.631

What Works Clearinghouse. (2020). *What Works Clearinghouse standards handbook, version 4.1.* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf

Wiltsey-Stirman, S., Baumann, A.A., & Miller, C.J. (2019). The FRAME: An expanded framework for reporting adaptations and modifications to evidence-based interventions. *Implementation Science, 14*, Article 58. https://doi.org/10.1186/s13012-019-0898-y

Table 1

Participant and Setting Characteristics

Teachers (N=104)	% Reported	Mean	Min	Max
Schools				
% Free or Reduced Lunch	39%	58%	1%	100%
% White	32%	43%	1%	88%
School Type	97%			
% Preschool	-	29%	-	-
% Elementary	-	60%	-	-
% Middle	-	11%	-	-
School Location (in US)	74%			
Northeast	-	48%	-	-
Northwest	-	13%	-	-
Midwest	-	4%	-	-
East	-	4%	-	-
South	-	13%	-	-
Southeast	-	8%	-	-
Southwest	-	8%	-	-
School Urbanicity	52%			
Urban	-	31%	-	-
Suburban	-	56%	-	-
Rural	-	13%	-	-
Teachers				
% Female	81%	99%	75%	100%
% White	68%	84%	0%	100%
Years experience	87%	10.36	0	26
Grade Taught	88%	-	-	-
Preschool/Head Start	-	30%	-	-
Kindergarten	-	12%	-	-
1st Grade	-	7%	-	-
2nd Grade	-	21%	-	-
3rd Grade	-	13%	-	-
4th Grade	-	4%	-	-
5th Grade	-	11%	-	-
6th Grade	-	0%	-	-
7th Grade	-	3%	-	-
8th Grade	-	1%	-	-

Table 2

Implementation Strategy Characteristics

Author and Year	Implementation Strategy Names (Actions)	Strategy Type	Quality Index	Strategy Temporality	Strategy Deliverer (Actor)
Bethune, 2017	Pre-coaching meeting, including planning, instruction, modeling; side-by-side coaching, including modeling, praise, and error correction; follow-up performance feedback session	Multi-faceted	4	AI	School-Based Instructional Coach
Collier-Meek et al., 2016	Direct training, action and coping planning, modeling, role-play, performance feedback	Blended	4	AI	Graduate Students
Dart et al., 2012	Test-driving interventions	Discrete	3	AI	Graduate Students
DiGennaro et al., 2005	Performance feedback with negative reinforcement via meeting cancelation, and dynamic fading	Multi-faceted	4	AI, MA	Consultant
Dufrene et al., 2012	Direct training, bug-in-ear prompting, performance feedback	Multi-faceted	4	AI	Graduate Students
Fullerton et al., 2009	Direct Training, performance feedback delivered via note or email, behavioral cues	Multi-faceted	3	AI	PBS Consultant
Gross et al., 2014	Performance feedback and directed rehearsal with 3 levels of fading	Multi-faceted	3	AI, MA	Consultant
Hemmeter et al., 2011	Direct Training and Email feedback	Multi-faceted	4	AI	Graduate Students
Hundert et al., 1992	Provide rationale for change, action plan, 30-minute meeting with supervisor	Discrete	5	PI	School Supervisor
Kleinert et al., 2017	Modified Classroom Check Up (CCU): Feedback session with data, menu of options, planning, and goal setting	Blended	3	AI	Researchers
McKenney et al., 2013	Training, modeling, performance feedback on mock FA	Multi-faceted	3	AI	Researchers

Myers et al., 2011	Tiered strategies: brief consultation, rationale and examples, contingent praise. Daily scripts, behavioral and environmental cues	Multi-faceted	4	AI	Researchers
Noell, et al., 2002	Data Review, Performance Feedback with Fading	Multi-faceted	4	AI, MA	Graduate Students
Oliver et al., 2015	Training, presenting content, modeling, observing video, practice with performance feedback, self-monitoring upon stable implementation	Multi-faceted	5	AI, MA	Researchers
Reinke et al., 2007	Pre-implementation group consultation and visual performance feedback delivered midway through experimental phase	Multi-faceted	3	PI, AI	Researchers
Reinke et al., 2008	Classroom Check-Up: Feedback sessions with data, provide menu of EBP, action planning, ongoing monitoring, goal setting.	Blended	3	AI	Researchers
Riley-Tillman & Eckert, 2001	School-based problem-solving consultation and coaching	Blended	4	AI	Researcher and Graduate Students
Rispoli et al., 2015	TBFA training (Pre and Post): training, role playing, immediate corrective feedback, follow-up feedback sessions.	Multi-faceted	4	AI	Researcher or Graduate Student
Rodriguez et al., 2009	Performance feedback with data, problem solving, planning	Multi-faceted	4	AI	Graduate Candidates
Sanetti et al., 2014	Consultation, Implementation Planning	Blended	4	AI	Graduate Students
Sanetti et al., 2015	Implementation Planning	Blended	5	AI	Graduate Students
Sanetti et al., 2013	Verbal and Graphic Performance Feedback	Discrete	4	AI	School Social Worker and Special Education Teacher
Sanetti & Collier-Meek, 2015	Tiered strategies: Direct Training (1) Implementation Planning (2). Modeling (3)	Multi-faceted	4	AI	Graduate Students
Simonsen et al., 2017	Targeted Professional Development	Multi-faceted	4	AI	Researcher
Smith et al., 2011	Consultation with direct training, practicing & performance feedback	Blended	4	AI	Researcher and Graduate Students
Stormont et al., 2007	Consultation with direct training, practicing & performance feedback	Blended	3	AI	Researcher and Graduate Students

Note. PI = Pre-Implementation Strategy, AI = Active-Implementation Strategy, MA = Maintenance Strategy

Table 3

Number of Studies and Average Effect Size for Each Behavior Change Technique

Behavior Change Technique	Number and Percent of Studies	*k*	Hedge's *g*	SE	Tau	SD
Performance Feedback	22 (79%)	82	2.29	.24	.73	.07
Procedural Knowledge	18 (64%)	71	2.84	.28	.89	.04
Action Planning	13 (46%)	51	2.41	.33	.82	.05
Directed Rehearsal / Practice	12 (43%)	42	2.74	.32	.97	.03
+/- Reinforcement (Praise / Meeting Cancelation)	11 (39%)	47	2.96	.37	.83	.05
Problem-Solving Barriers / Coping Planning	8 (29%)	32	1.99	.31	.72	.06
Goal Setting	7 (25%)	28	2.74	.39	.83	.05
Modeling	7 (25%)	24	2.67	.50	.85	.05
Review Student Data	6 (21%)	23	2.17	.42	.72	.07
Role-Play	5 (18%)	18	2.37	.33	.90	.04
Video Modeling	5 (18%)	18	3.70	.72	.92	.04
Discuss Positives to Implementation / Rationale	5 (18%)	18	2.25	.48	.82	.06
Prompts	5 (18%)	16	2.64	.71	.80	.09
Adapt Intervention to fit Context	4 (14%)	14	2.72	.42	.87	.02
Self-Monitoring	4 (14%)	17	2.64	.89	.80	.09
Self-Evaluation	2 (7%)	9	3.06	.59	.86	.13
Fading of Supports*	3 (11%)	9	0.38	.08	--	--

Note. *k* = number of time-series graphs. *Small effect may be caused by the fact that this was compared to treatment and was employed after active supports were removed. Effect sizes averaged over studies and active ingredient using a random effects model and inverse variance weights to produce the average effect size. Systematic fading was not assessed in overlap effect-size calculations. Not all active ingredients are included; Only those that appeared across two or more studies were quantified.

Table 4

Number of Studies and Average Effect Size for Each Mechanism of Action

Mechanism of Action (MoA)	**Number and Percent of Studies**	***k***	**Hedge's *g***	***g* SE**	**Tau**	**Tau SD**
Knowledge (Kn)	18 (64%)	71	2.84	.28	.79	.04
Skill (Sk)	22 (79%)	85	2.73	.26	.85	.03
Beliefs about Capabilities (BaCa)	23 (82%)	92	2.77	.25	.83	.04
Beliefs about Consequences (BaCo)	9 (32%)	35	2.24	.38	.76	.06
Reinforcement (Re)	10 (36%)	53	2.84	.43	.85	.05
Intentions (In)	5 (17%)	21	2.44	.30	.83	.06
Goals (Go)	7 (25%)	28	2.74	.34	.83	.05
Memory, Attention & Decision Processes (MADP)	5 (18%)	16	3.35	.71	.79	.09
Environmental Context and Resources (ECR)	15 (53%)	67	2.27	.30	.81	.04
Social Influences (SI)	6 (24%)	23	2.97	.40	.89	.04
Behavior Regulation (BR)	10 (36%)	42	2.24	.32	.77	.06
Attention Toward the Behavior (Attb)	6 (21%)	22	2.44	.50	.84	.05
Motivation (Mo)	22 (78%)	82	2.29	.24	.73	.07
Feedback Processes (FP)	23 (82%)	91	2.35	.25	.76	.06
Social Learning / Imitation (SLI)	11 (39%)	38	2.92	.40	.88	.04
Behavioral Cueing (BC)	16 (57%)	59	2.45	.30	.81	.04

Note. Hedge's *g* calculated by including only those participants where mechanism was present. Effect sizes averaged over studies and active ingredient using a random effects model and inverse variance weights to produce the average effect size. *k* = number of time-series graphs that incorporated the mechanism of action.

Table 5

Table 5

Evidence-Based SEB Practice Being Implemented

Author & Year	SEB EBP Being Implemented	Intervention Tier
Bethune et al., 2017	PBIS Practices	1
Carter et al., 2010	PBS Practices	1
Collier-Meek et al., 2016	Classroom Management	1
Collier-Meek et al., 2017	Good Behavior Game or Caught Being Good Game	1
Dart et al., 2012	Response Cost, Behavior Specific Praise, Self-Monitoring	2
DiGennaro et al., 2005	Behavior Specific Praise	2
Dufrene et al., 2012	Praise Statements	1
Fullerton et al., 2009	Behavior Specific Praise	2
Gross et al., 2014	Modified Check-In Check-Out	2
Hemmeter et al., 2011	Behavior Specific Praise	1
Hundert & Hopkins, 1992	Directing Attention Toward Students with Disabilities in Class	2
Kleinert et al., 2017	Opportunities to Respond and Praise Statements	1
McKenney et al., 2013	Functional Analysis Procedure	3
Myers et al., 2011	Behavior Specific Praise	1
Noell et al., 2002	Individualized Behavior Management Intervention	3
Oliver et al., 2015	Good Behavior Game	1
Reinke et al., 2007	Behavior Specific Praise	1
Reinke et al., 2008	Praise Statements	1
Riley-Tillman & Eckert, 2001	Praise Statements	1
Rispoli et al., 2015	Trial-Based Functional Analysis	3
Rodriguez et al., 2009	First Steps to Success	2
Sanetti & Collier-Meek, 2015	Behavior Support Plans	3
Sanetti et al., 2013	Classroom Management Plan	2
Sanetti et al., 2014	Behavior Support Plans	3
Sanetti et al., 2015	Pax Good Behavior Game	1
Simonsen et al., 2017	Behavior Specific Praise	1
Smith et al., 2011	Pre-corrective Statements and Behavior Specific Praise	1
Stormont et al., 2007	Behavior Specific Praise	2

Note. This table represents a crosswalk of the social-emotional or behavioral evidence-based practice being implemented by the teachers in each study and the intervention tier at which it was delivered.

Table 6

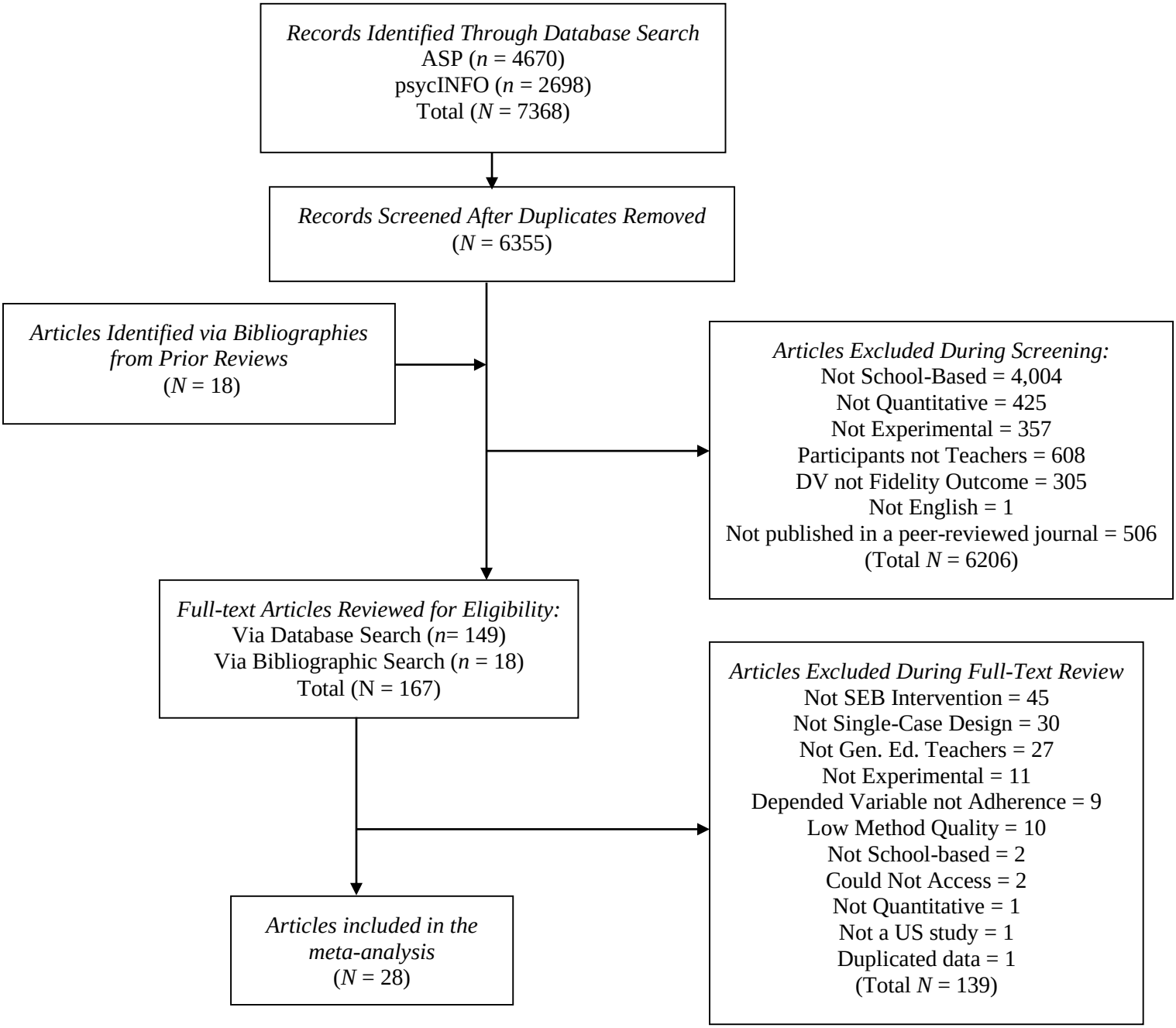
Random Effects Omnibus Moderator Model Output

	Estimate	StdErr	Z-value	*p*-value	95% CI
X- intercept	2.098	0.125	16.853	0.000	[1.854, 2.342]
Grade Taught	-0.114	0.131	-0.876	0.381	[-0.370, 0.141]
Years Experience	0.073	0.125	0.582	0.561	[-0.172, 0.318]
Intervention Tier	-0.112	0.122	-0.904	0.366	[-0.351, 0.129]
Quality Indicator	-0.263	0.134	-1.874	0.061	[-0.537, 0.012]
# BCTs	0.490	0.129	3.797	0.000	[0.237, 0.743]

Note: All values are standardized. Number of studies = 28; Number of outcomes = 122; Qe(df = 117) = 466.48, $p < .001$; Qm(df = 5) = 24.0076, $p = .0002$; $\tau^2 = 1.31$; $I^2 = 79.09\%$; $R^2 = 20.45\%$.

Figure 1

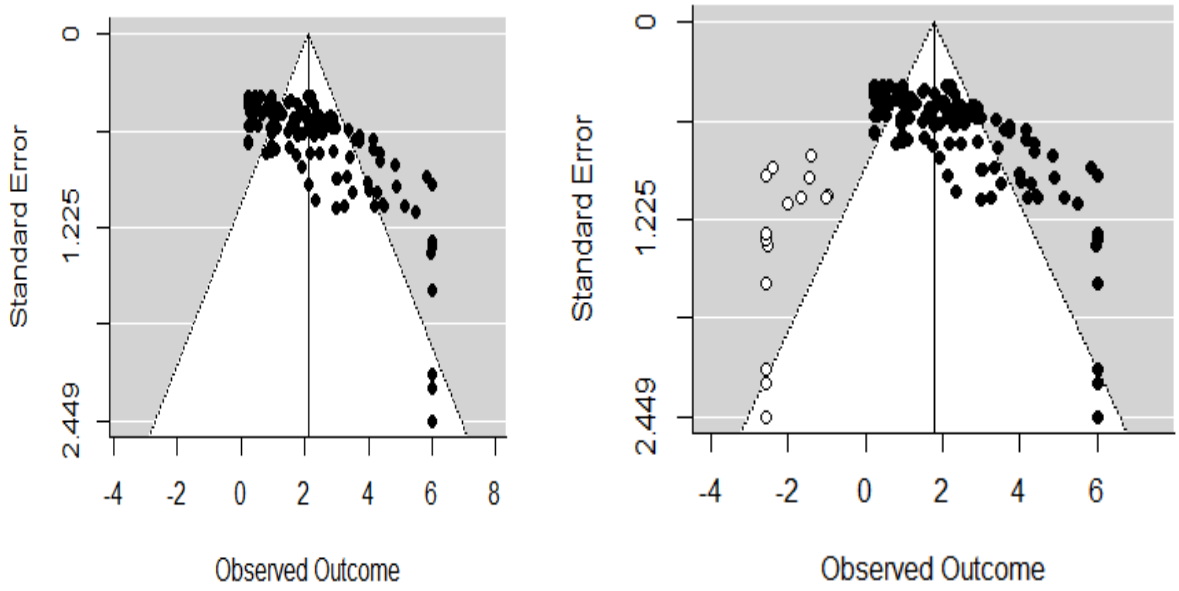
PRISMA Flow Chart



Note. This diagram depicts the flow of information through the different phases of the systematic review. It maps out the number and source of records identified, included and excluded, and the reasons for exclusions.

Figure 2

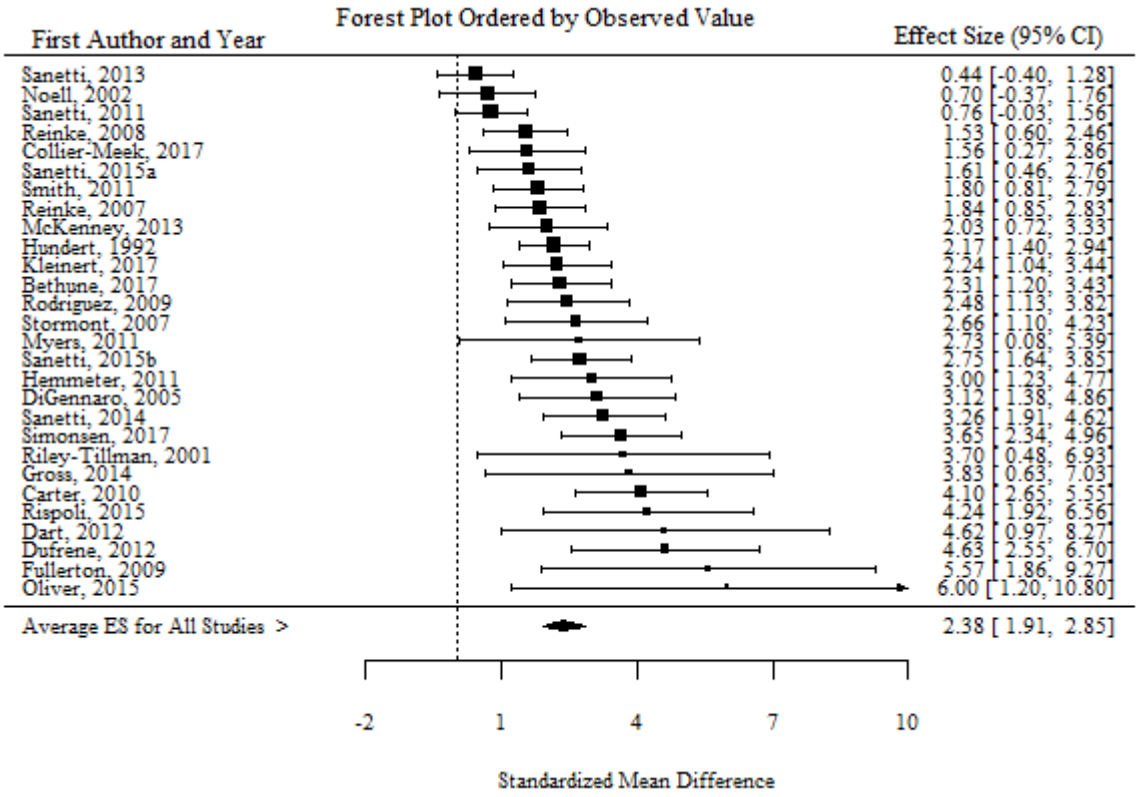
Funnel Plots



Note. These funnel plots represent the distribution of effect sizes without the trim-and-fill method (left), and a funnel plot including the trim-and-fill method (right). Black dots represent observed datapoints, while white dots represent studies that should be present if publication bias was not observed in the dataset.

Figure 3

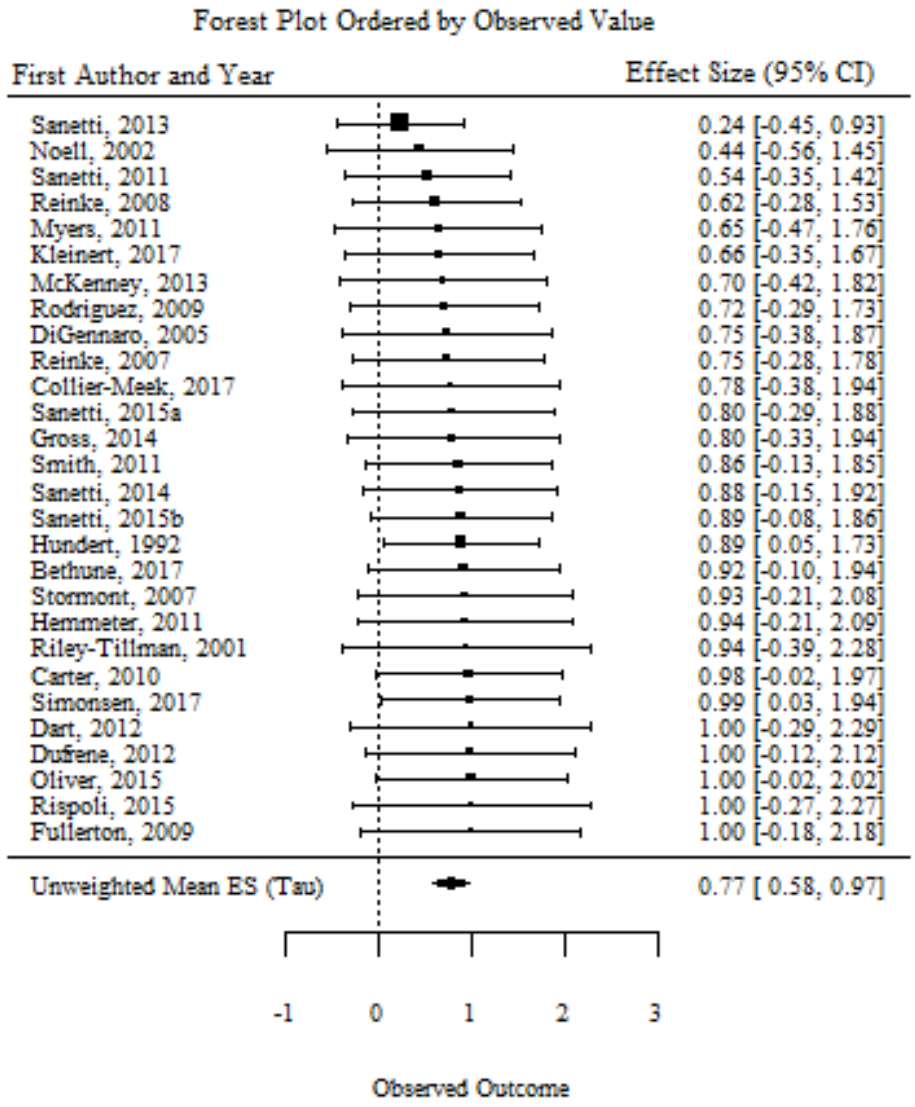
Forest Plot of Hedge's g Effect Sizes for Each Study



Note. Larger dots represent a greater precision (i.e. lower variance among participants within the study). Horizontal bands indicate the standard error of each study estimate. These effect sizes represent unweighted averages obtained to make the plot more readable. The diamond shape at the bottom represents the average effect size across studies.

Figure 4

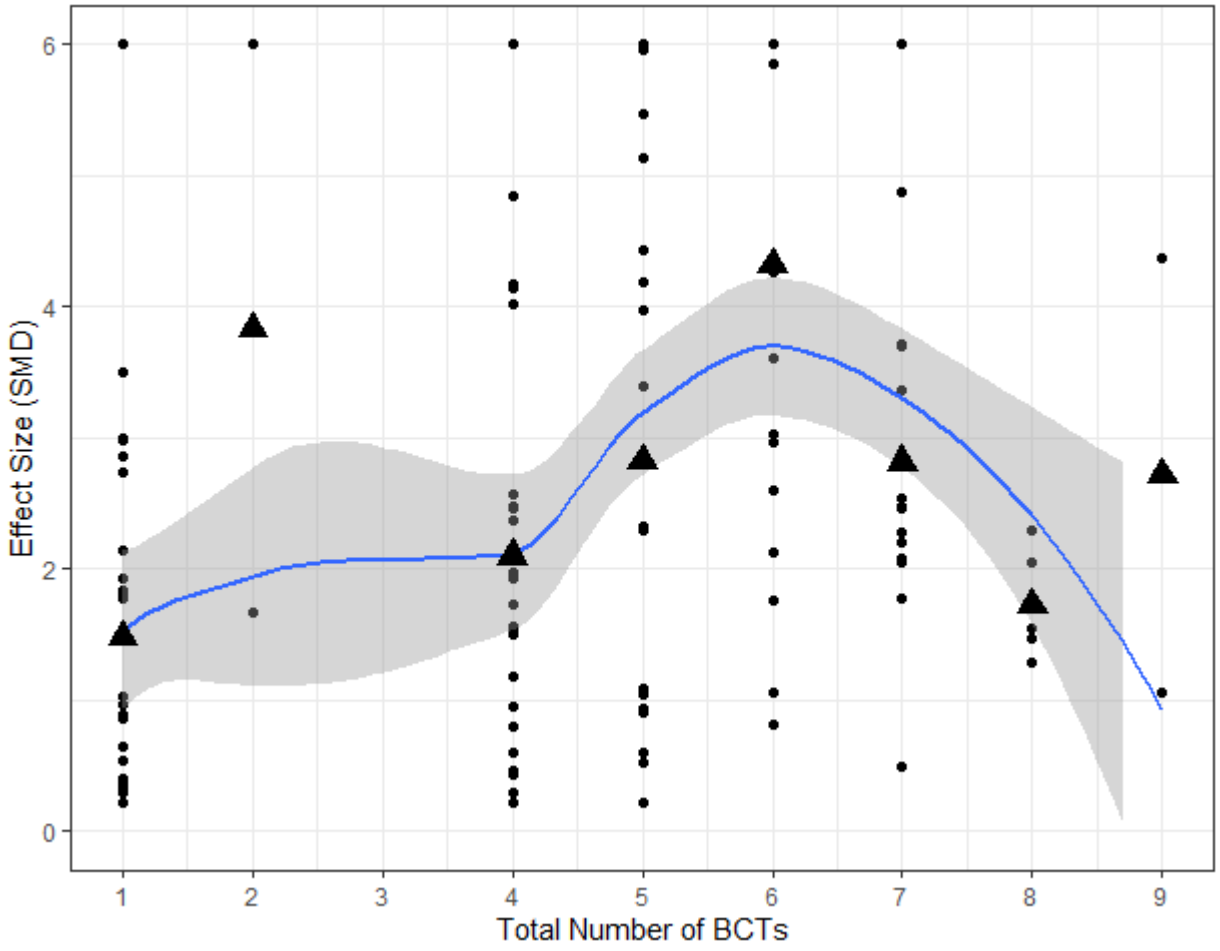
Forest Plot of Tau Effect Sizes for Each Study



Note. Effect size limited to 1, which is the upper limit of Tau. Larger dots represent a greater precision (i.e. lower variance among participants within the study). Horizontal bands indicate the standard error of each study estimate. These effect sizes represent unweighted averages obtained to make the plot more readable. The diamond shape at the bottom represents the average effect size across studies.

Figure 5

Relationship of BCTs to Effect Size (Hedge's g)



Note. This plot depicts the total number of BCTs utilized for each participant by their effect size. The loess smoother indicates a nonlinear relationship between effect size and the total number of BCTs used, with a diminishing trend.

Supplemental Table 1

Codebook Guide with Operational Definitions of Key Variables and Item Reliabilities

e.g. 1. Code (*operational definition of the code):*	*IRR of Categorical Variables*	
	% agreement	Kappa
TOTAL averages:	93%	.74
1. Study ID #:		
2. First author last name (*Name of publishing author)*:		
3. Publication year (*Year study was published)*:		
4. Journal (*Journal study was published in)*:		
5. Study Source (*Database search or citation search):*		
Setting:		
1. School Grades (*pre-K–K; elementary (grades 1–5); middle (grades 6–8); high schools (grades 9–12)):*	90%	.674
2. School Demographic Breakdown		
a. race/ethnicity of school: (*schoolwide percent of white/Caucasian students):*	Sparse	
b. Urbanicity: (*urban, suburban, rural*):	Sparse	
3. School Socioeconomic Status (*School level % Free or Reduced-Price Lunch*):	Sparse	
4. School Geographic Region (*i.e. where in the US is the school located):*	Sparse	
Participants		
1. Teacher race/ethnicity (*Race or ethnicity of the teachers implementing the SEB EBP):*	100%	1
2. Teacher sex/gender (*Sex or gender of the teachers implementing the SEB EBP)*:	100%	1
3. Teacher experience (*Years spent teaching*):	100%	1
Study Characteristics		
1. Single-Case Study Design: (*i.e. ABAB / Multiple baseline design)*	100%	1
2. Number of phases / replications of effect: (*number of phases in the single-case design)*	100%	1
3. Number of data points per phase	100%	1
Independent Variables		
1. Implementation strategy being delivering to teacher *(the interventions that target educator behavior change to facilitate the uptake and delivery of EBPs)*		
a. Name of strategy: (*name of the implementation strategy(ies) provided by actor(s))*		

b. Type of strategy *(**discrete** = 1 strategy, **multifaceted** = 2 or more strategies, **blended** = 2 or more strategies delivered in a manualized program)*	90%	.83
c. Temporality *(**pre-implementation** = occurs prior to teacher beginning to implement the SEB practice; **active-implementation =** occurs during implementation; **maintenance strategy** = occurs after implementation in order to sustain)*	80%	.60
d. Name and Number of Active Ingredients (*unique implementation strategies used; i.e. planning, coaching, performance feedback, motivational intervention, teaching intervention, etc).*	89%	.78
ii. Procedural knowledge	81%	0
iii. Performance feedback	100%	1
iv. Action Planning	81%	.63
v. Directed Rehearsal / Practice	100%	1
vi. Reinforcement	80%	.61
vii. Problem solving	92%	.81
viii. Goal setting	91%	.67
ix. Modeling	100%	1
x. Adapt intervention	81%	0
xi. Role Play	81%	.63
xii. Video Modeling	100%	0
xiii. Prompts	81%	0
a. Mechanism of Action (*the processes or events through which an implementation strategy operates to effect desired implementation outcomes as determined by the Theoretical domains framework, and Theory and Techniques Tool)* https://theoryandtechniquetool.humanbehaviourchange.org/tool	88.64%	.77
i. Kn	81%	0
ii. Sk	84%	.70
iii. BaCa	88%	.77
iv. BaCo	100%	1
v. Re	100%	1
vi. In	100%	1
vii. Go	100%	1
viii. MADP	100%	1
ix. ECR	100%	1

x. SI	100%	1
xi. BR	81%	.63
xii. Attb	100%	1
xiii. Mo	86%	0
xiv. FP	86%	0
xv. SLI	86%	.82
xvi. BC	81%	.63
a. Implementation fidelity (*fidelity with which the implementation strategy was delivered to the teacher)*	90%	.80
2. Social-emotional behavioral (SEB) evidence-based practice(s) the teacher is implementing: (*SEB defined as any non-academic intervention that involves any of the following: (1) how students interact with others and impacts the quality of their relationships (i.e., social), (2) the difficulties students have regulating and managing their feelings (i.e., emotions), and (3) the behaviors students exhibit that are disruptive to learning environments (i.e., behavior)*	100%	1
a. Level of service (*is the SEB EBP the participating teacher is delivering being targeted toward an individual student (**tier 3**), a small group (**tier 2**), or class-wide (**tier 1**)*):	90%	.59
3. Actor (*individual(s) delivering the implementation strategy*): (Outside Researcher or School-based staff)	100%	1
a. Outside expert/consultant or internal school personnel (if other, describe)	100%	1
i. Role, if school personnel (i.e. *school psychologist, principal, instructional coach; graduate student, faculty member)*	100%	1
ii. Training (*hours of training to deliver implementation strategies / consultation received*)	Sparse	
Dependent Variables		
1. Teacher Outcome Variable(s) (*i.e. **adherence**: whether the program service or intervention is being delivered as it was designed; % steps completed / praise statement or reprimand frequency)*	90%	.80
2. Method of data collection (*Direct observation or Permanent product review*)	100%	1
3. Interrater relatability (*Kappa and % agreement of DV)*	N/A	N/A
Social Validity		
1. Social Validity Scale(s) used and Avg. Score / SD (*i.e. feasibility, acceptability, or appropriateness)*	Sparse	

Note. Some variables were too sparse to be included in reliability.

Supplemental Table 2

Behavior Change Technique Implementation Strategy Definitions

Behavior Change Technique	Definition
Directed Rehearsal	Practicing delivery of the intervention with error correction and correcting missed steps of an intervention with a consultant
Goal Setting	Determining a desired targeted outcome
Action Planning	Creating a detailed plan outlining actions needed to reach one or more goals.
Coping Planning	A process outlining and problem-solving potential barriers to implementation.
Modeling	A coach or consultant performing the desired behavior while consultee observes.
Role Play	A consultant and consultee take turns performing and receiving the intervention
Performance Feedback	Consultant and consultee meet and discuss implementation integrity data
Review Student Data	Review data tracking student behavior
Prompting	A process where a consultee is cued or reminded to perform a behavior (i.e. email, post-it notes, verbally)
Video Modeling	Consultee views video recording of individual implementing desired behavior / intervention
Self-Monitoring	Consultee monitors their own implementation behavior
Self-Evaluating	Reviewing one's own performance to determine whether a goal was met.
Fading	A process where supports are systematically withdrawn
Discuss Positives to Implementation	Consultant discusses the importance of implementation to consultee by tying adherence to outcomes

Supplemental Table 3

Mechanisms of Action and Definition

Mechanism of Action	**Definition**
Knowledge (Kn)	An awareness of the existence of something
Skills (Sk)	An ability or proficiency acquired through practice
Social Professional Role and Identity (SPRI)	A coherent set of behaviors and displayed personal qualities of an individuals in a social or work setting
Beliefs about Capabilities (BaCa)	Beliefs about one's ability to successfully carry out a behavior
Optimism (Op)	Confidence that things will work out for the best and that desired goals will be attained
Beliefs about Consequences (BaCo)	Beliefs about the consequences of a behavior (what will be achieved or lost by carrying out the behavior).
Reinforcement (Re)	Process by which the frequency or probability of a response is increased through a dependent relationship or contingency with a stimulus or circumstance
Intentions (In)	A conscious decision to perform a behavior or resolve to act in a certain way
Goals (Go)	Mental representations of outcomes or end states than an individual wants to achieve
Memory, Attention, & Decision Processes (MADP)	Ability to retain information, focus on aspects of the environment and choose between two or more alternatives
Environmental Context & Resources (ECR)	Aspects of a person's situation or environment that discourage or encourage the behavior.
Social Influences (SI)	Those interpersonal processes that can cause oneself to change one's thoughts, feelings, or behaviors
Emotion (Em)	A complex reaction pattern involving experiential, behavioral, and physiological elements

Behavioral Regulation (BR)	Behavioral, cognitive and/or emotional skills for managing or changing behavior
Norms (No)	The attitudes held and behaviors exhibited by other people within a social group
Subjective Norms (SN)	One's perceptions of what most other people within a social group believe and do
Attitudes toward the behavior (Attb)	The general evaluations of the behavior on a scale ranging from negative to positive
Motivation (Mo)	Processes relating to the impetus that gives purpose or direction to behavior
Self-image (Si)	One's conception and evaluation of oneself, including psychological and physical characteristics, qualities, and skills
Needs (Ne)	Deficit of something required for survival, well-being or personal fulfilment
Values (Va)	Moral, social, or aesthetic principles accepted by an individual or society as a guide to what is good, desirable or important
Feedback Processes (FP)	Processes through which current behavior is compared against a particular standard
Social Learning / Imitation (SLI)	A process by which thoughts, feelings, and motivational states observed in others are internalized and replicated without the need for conscious awareness
Behavioral Cueing (BC)	Processes by which behavior is triggered from either the external environment, the performance of another behavior, or from ideas appearing in consciousness
General Attitudes / Beliefs (GAB)	Evaluations of an object, person, group, issue or concept
Perceived Susceptibility/Vulnerability (PSV)	Perceptions of the likelihood that one is vulnerable to a threat