ORIGINAL PAPER

Longitudinal Effects of a Motivationally Focused Strategy to Increase the Yield of Training and Consultation on Teachers' Adoption and Fidelity of a Universal Program

James L. Merle[1] · Clayton R. Cook[2] · Michael D. Pullmann[3] · Madeline F. Larson[2] · Corinne M. Hamlin[2] · Maria L. Hugh[3] · Stephanie K. Brewer[3] · Mylien T. Duong[3] · Mahasweta Bose[2] · Aaron R. Lyon[3]

Abstract

Group-based didactic training is a cornerstone implementation strategy used to support the adoption and delivery of evidence-based prevention programs (EBPP) by teachers in schools, but it is often insufficient to drive successful implementation. Beliefs and Attitudes for Successful Implementation in Schools for Teachers (BASIS-T) is a theory-based, motivational implementation strategy designed to increase the yield of EBPP training and consultation. The purpose of this study was to examine the longitudinal effects of BASIS-T on hypothesized mechanisms of behavior change (e.g., attitudes toward EBPP, self-efficacy, intentions to implement) and implementation and student outcomes associated with a well-established universal prevention program—the good behavior game (GBG). This pilot trial included 82 elementary school teachers from nine public elementary schools who were randomly assigned at the school-level to the BASIS-T ($n = 43$) or active comparison ($n = 39$) condition, with both conditions receiving training and consultation of the good behavior game by a third-party purveyor. Analyses included mixed-effects and multilevel growth modeling of adoption, mechanisms of behavior change, and student behavior outcomes. Meaningful effects were found favoring BASIS-T on immediate adoption of the GBG within the first month of school (74% vs. 40%) and self-efficacy ($p < 0.05$). These findings advance our understanding of the type of implementation strategies that complement pre-implementation training and post-training consultation in schools by identifying the importance of task self-efficacy as a mechanism of behavior change related to adoption for prevention programming.

Keywords Implementation · Implementation strategy · Training · Motivational interviewing · School · Behavior change · Mechanism

Over the last three decades, numerous EBPPs have been designed for teachers to support students in schools (Sanetti & Collier-Meek, 2019; Tanner-Smith et al., 2018). However, there is a significant implementation gap resulting in limited uptake and routine use of EBPPs (Lyon & Bruns, 2019), thus limiting their reach and overall impact. Staff development through didactic training is a core implementation strategy to address implementation gaps (Lyon et al., 2017; Owens et al., 2014). Pre-implementation didactic training, however, is often insufficient or suboptimal at promoting teachers' EBPP adoption (i.e., initial uptake) and fidelity (i.e., adherence to core practices) in pursuit of achieving meaningful changes in youth outcomes (Joyce & Showers, 2002; Merle et al., 2022a, 2022b; Reinke et al., 2008). Only around 40% of teachers adopt a new practice with training alone (Sanetti et al., 2013), and among those who do adopt, fidelity deteriorates within 10 days of adoption (Reinke et al.., 2008; Sanetti & Kratochwill, 2009). Many implementation efforts begin with the provision of training and do not incorporate the individual motivational factors linked to implementer behavior change (Beidas & Kendall, 2010). Because behavior change is necessary for successful implementation (Lyon et al., 2018), failing to address individual-level motivational

✉ James L. Merle
james.merle@utah.edu

1   Department of Population Health Sciences, University of Utah School of Medicine, 295 Chipeta Way, Salt Lake City, UT 84108, USA

2   University of Minnesota, 56 E River Pkwy, Minneapolis, MN 55455, USA

3   University of Washington, 6200 NE 74th St, Suite 100, Seattle, WA 98115, USA

and volitional factors may moderate implementation outcomes associated with training and follow-up consultation. Given this, our team developed an implementation strategy to complement EBPP training and consultation that targets individual-level mechanisms of behavior change. In this pilot randomized controlled trial, we investigated the longitudinal effects of Beliefs and Attitudes for Successful Implementation in Schools for Teachers (BASIS-T) against an active comparison (AC) control condition on mechanisms of behavior change and implementation outcomes (i.e., adoption and fidelity) across one year within the context of training and consultation on a well-established universal EBPP.

Implementation Strategies to Support Teachers

In order to ensure that teachers successfully implement EBPPs, they need to be supported both prior to and during implementation (Merle et al., 2022a, 2022b). The discrete practices used to support teacher implementation, *implementation strategies*, represent the methods and techniques used to promote implementation outcomes (e.g., appropriateness, feasibility, adoption and fidelity; Powell et al., 2015). School-based implementation research has largely focused on pre-implementation training, and active-implementation coaching and consultation supports (Fallon et al., 2015; Noell et al., 2014; Solomon et al., 2012; Stormont et al., 2015). Traditional training and consultation models are cornerstone strategies designed to support implementation (Beidas et al., 2012), yet many implementers fail to successfully deliver an EBPP after receiving training and consultation, even when these supports are high quality (Nadeem et al., 2018). In fact, many practitioners are ambivalent to change or lack the self-efficacy to overcome perceived barriers to implementation, which influences adoption decisions (Dart et al., 2012). Moreover, early intervention adopters tend to stop using a new practice shortly after starting (e.g., within 4–6 weeks; Stirman et al., 2012).

In response to these implementation challenges, researchers have developed and tested the impact of specific consultative strategies, such as performance-based feedback (Collier-Meek et al., 2017), prompts/reminders, (Fallon et al., 2018) and practice monitoring (Mouzakitis et al., 2015), on intervention fidelity. Sanetti and colleagues (2013) also utilized the Health Action Process Approach (HAPA) to develop an implementation planning strategy to boost implementation behaviors. However, these implementation strategies often assume that implementers are motivated to adopt and deliver the EBPP (Reinke et al., 2011). Additionally, follow-up implementation supports require additional resources, such as staff and time. Robust pre-implementation experiences may have the potential to yield better early

adoption results, which could reduce the number of teachers needing intensive follow-up supports and thus reducing the resource burden for already resource-limited schools. Finally, although a variety of strategies have been developed and tested with teachers, a lack of strategy specificity in the literature has made it difficult to discern *how* and *why* these strategies achieve their effect (Merle et al., 2022a, 2022b).

Gaps in the Research

This study seeks to address multiple gaps in the literature. First, there are few pre-implementation strategies available for use in schools that are informed by behavior change theory. Therefore, there is a continued need for school-based implementation research that utilizes theory to develop implementation strategies (Larson et al., 2021; Lyon et al., 2019; Wolfenden et al., 2021). Since the yield of training alone results in around 40% of participants implementing a new practice (Sanetti et al., 2013), there is a need to improve upon the techniques used to enhance training, such as ensuring that participants are motivated prior to beginning training, and have a plan to enact the practice prior to beginning implementation. Increasing the yield of training will theoretically reduce the number of teachers needing intensive, ongoing support. There is also a limited understanding of precisely how implementation strategies achieve their effects, which has created a "black box" problem in implementation research (Lewis et al., 2018). Moreover, there are few studies that explicitly target the underlying mechanisms of behavior change within school-based implementation. This study will seek to build off previous work to uncover the behavior change mechanisms that are most important to support teacher EBPP implementation in schools.

Beliefs and Attitudes for Successful Implementation in Schools (BASIS)–for Teachers

BASIS was developed as an implementation strategy to complement the training and consultation provided by EBP purveyors. It is designed to target individual-level mechanisms of behavior change. BASIS is grounded in two well-established theories of adult behavior change—the Theory of Planned Behavior (TPB; Ajzen, 1985) and Health Action Process Approach (HAPA; Schwarzer, 2008, Schwarzer et al., 2011). Together, HAPA and TPB outline: (1) *motivational factors* (i.e., task self-efficacy, subjective norms, attitudes/outcome expectancies) that act as hypothesized determinants of behavioral intentions (defined as an individual's commitment to exhibiting a particular set of behaviors*; Ajzen & Manstead, 2007) and (2) *volitional factors* (Schwarzer et al., 2011) that involve mechanisms and supports that are essential to enable individuals to enact the

behaviors they intend to perform (Hagger & Chatzisarantis, 2014).

As most pre-implementation didactic trainings for teachers occur in a large group, BASIS-T also uses a group-based format, which increases feasibility and reduces cost (Joyce & Showers, 2002). Theoretically, BASIS-T works by shifting teachers' motivational mechanisms to implement by impacting attitudes, self-efficacy, and social norms during the pre-session—and in the post-EBPP session, it is designed to increase volitional mechanisms, and intentions to implement through group-based implementation planning (Ajzen, 1985; Schwarzer et al., 2011). Together, BASIS-T aims to promote early adoption and persistence towards reaching high fidelity. Moreover, to protect against deterioration of intervention fidelity, teachers receive a digitally delivered booster 15 days post-training. BASIS-T sits at the intersection between the preparation and initial implementation phases of the implementation process to promote greater numbers of teachers who initially adopt the EBPP in response to training and higher levels of fidelity among teachers in response to follow-up consultation (Aarons et al., 2011). BASIS was developed iteratively over time originating from practiced-based evidence, and it has been tested with different school-based professionals (Cook et al., 2015; Larson et al., 2021; Lyon et al., 2019). Detailed descriptions of BASIS-T content, structure, and specificity can be found in the Procedures section of this paper, Supplemental Files 1 and 3, and in Larson et al. (2021).

In the first pilot, we examined a preliminary version of BASIS using a pre-post, no control group design, with educators from 62 schools that bookended training in school-wide positive behavior intervention and supports (SW-PBIS). Pre-post-training surveys showed increased favorable attitudes toward EBPPs at post-intervention (Cohen's $d = 1.03$; Cook et al., 2015). Attitudes, in turn, were associated with two measures of EBPP fidelity (Cohen's $d = 0.51$; Cohen's $d = 0.67$). Our team also developed a version of BASIS for school-based clinicians (BASIS-C), and the results of a small-scale randomized trial supported its feasibility, acceptability, appropriateness, and its preliminary impact (Lyon et al., 2019), encouraging support for the proposed trial.

Most recently, our research team evaluated the initial mechanistic effects of BASIS-T on the initial impacts immediately following the post-training session for the current study (Larson et al., 2021). Results from pre-post-training revealed that, compared to the control group, the BASIS-T group had significantly higher self-efficacy ($p < 0.001$) and outcome expectancies ($p < 0.05$). Because this 2021 study only assessed the impact on mechanisms before implementation occurred, this study builds on that work through a longitudinal evaluation of (a) whether BASIS-T maintains effects on mechanisms of behavior change over the course of a school year and (b) the impact of BASIS-T on teachers' EBPP implementation outcomes (i.e., adoption and fidelity) and student behavioral outcomes in the context of purveyor provided training and consultation.

Purpose of Study and Research Questions

The purpose of this study was to conduct a pilot randomized controlled trial using a hybrid type III approach (Landes et al., 2019) to examine the longitudinal effects (one school year) of BASIS-T over and above EBPP training and consultation on (a) theorized mechanisms of behavior change, (b) teacher adoption and fidelity, and (c) classroom-level behavioral outcomes. Specifically, BASIS-T was deployed as an adjunct to "gold standard" training and consultation in the GBG, a well-established universal prevention program with evidence from several randomized trials supporting its effects on short- and long-term child outcomes (e.g., Bowman-Perrott et al., 2016). The following research questions guided this study:

1. Relative to the AC group, do teachers in the BASIS-T group have higher scores post-training on the hypothesized proximal mechanisms of change: attitudes, perceived social norms, self-efficacy, and intentions to implement? Do these scores sustain over the course of one school year?
2. Relative to the AC group, are teachers in the BASIS-T condition more likely to adopt the EBPP? Do they implement the EBPP with greater fidelity and frequency?
3. Relative to the AC group, are teachers in the BASIS-T condition associated with more positive classroom-level behavioral outcomes as indicated by the modified Direct Behavioral Rating (DBR) scale.
4. For significant mechanisms, is there preliminary evidence that the mechanism is associated with the effects of BASIS-T and with classroom-level behavioral outcomes?

Method

This study employed a hybrid type III cluster randomized controlled trial. The hybrid type III approach places greater emphasis on examining the implementation outcomes associated with implementation strategies, while also observing the intervention outcomes secondarily (i.e., the SEB outcomes associated with receiving the GBG; Landes et al., 2019).

Participants

Teachers from nine schools from a school district in the Northern Midwest region of the United States were recruited to participate. The partnering district was interested in collaborating in this study to improve the delivery of classroom-based practices to prevent social, emotional and behavioral (SEB) problems interfering with learning. Teachers were recruited based on their interest in learning about evidence-based classroom practices to prevent and address student SEB problems that interfere with learning. The schools had an average enrollment of 422 students (range 323–520), served a relatively racially (persons of color; $M = 26\%$, min.: 44%-96%, max.) and socioeconomically (free and reduced-priced lunch; $M = 44\%$; min.: 14%, max: 79%) diverse student population.

All nine elementary schools were actively implementing the universal level of school-wide positive behavior interventions and supports, although no data were available on the length of time it was being implemented nor the extent to which it was being implemented with fidelity. Randomization to condition occurred at the school level to reduce contamination across participants and was stratified to balance school characteristics (see Procedures). Out of 129 eligible staff in the nine schools, 88 teachers consented to participate (see CONSORT diagram; Fig. 1). Of these, 82 (93.1%) teachers attended the EBPP training. Of those attending the training, 81 teachers completed both pre- and post-training surveys, for a 98.7% retention rate. The teacher not completing post-training survey was in the BASIS-T condition. Chi-square and t-test analyses uncovered no statistically significant differences between post-survey completers and non-completers on gender, race, grade(s) taught, or any of the outcome variables collected at baseline. Table 1 displays participant demographics for the complete sample and each condition.

The Good Behavior Game (GBG) Training and Consultation

The GBG is a universal EBPP that uses an interdependent group contingency in which all members of a group (i.e., team) have access to the same consequence, based on the behavior of the collective group (Barrish et al., 1969). The GBG encourages teacher use of social learning principles within a game-like context to reduce disruptive behavior and
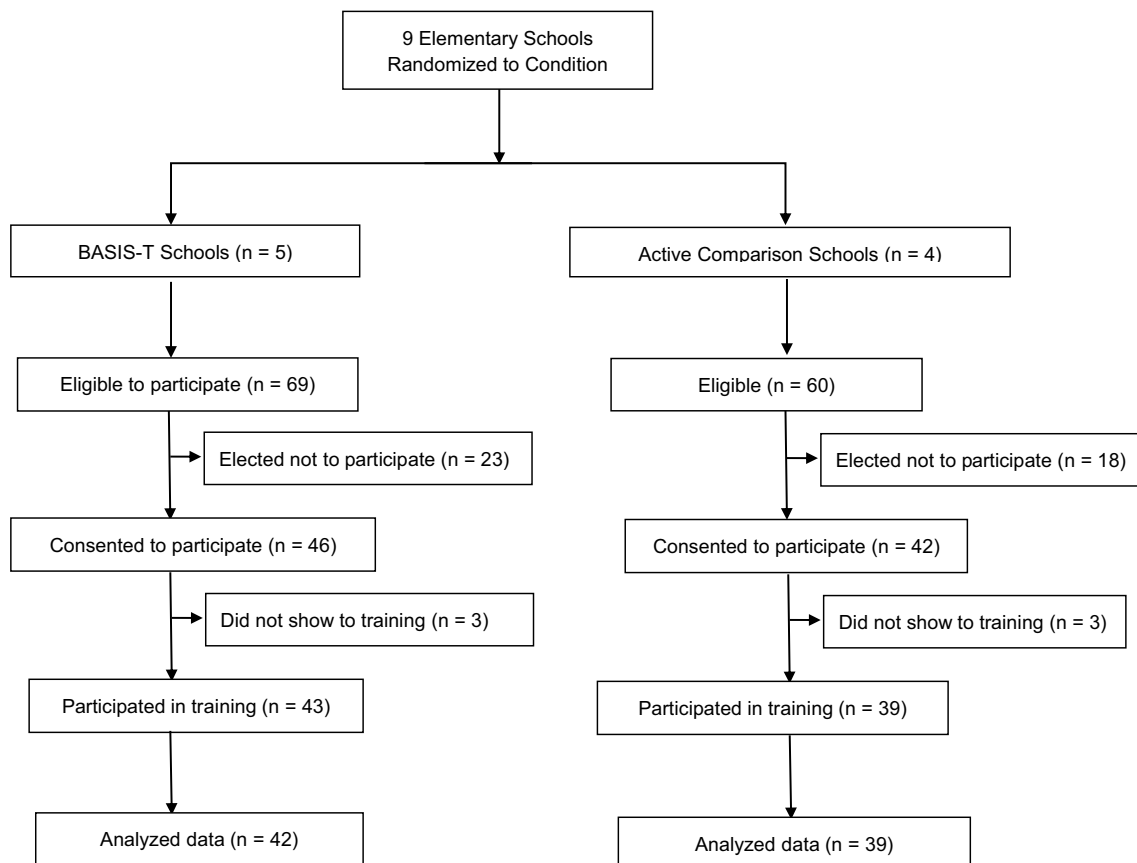

Fig. 1 CONSORT diagram for study participation

facilitate engagement. GBG has been evaluated for almost 50 years (e.g., Domitrovich et al., 2010) and endorsed as effective by numerous agencies (e.g., U.S. Center for Substance Abuse, NIDA, OJJDP), leading to its identification as a best practice. In the current study, all teachers participated in a standard, 1.5-day GBG training delivered by certified trainers, blinded to condition, after receiving BASIS-T or AC. GBG training included best practices for educational meetings: didactic content delivery, modeling, rehearsal activities, and performance-based feedback. Trainers also provided three follow-up consultation sessions across the academic year to embedded coaches in each school. The coaches served as a resource to teachers in their building who may need additional support to adopt and deliver GBG with fidelity.

Study Conditions

BASIS-T Condition Pre- and Post-Training Descriptions

Throughout BASIS-T, four components are embedded into a mixture of didactic information and interactive group-based activities, designed to target four theorized mechanisms of behavior change that lead to improved implementation outcomes (e.g., adoption and fidelity). See Supplemental File 1 for an in-depth description of BASIS-T theory of change, structure, components, and content as well as Larson et al. (2021). Supplemental File 3 displays the implementation strategy specifications in alignment with reporting guidelines (Proctor et al., 2013). The BASIS-T condition consisted of pre- and post-training sessions to bookend the GBG training. In this condition, teachers participated in a group-based, interactive, motivational session delivered by a member of the research team, which included a pre-training session prior to the GBG training (3 h). The pre-training experience involved a collection of group-based activities adhering to an *elicit-provide-elicit* structure, which included eliciting reflections from participants, providing information to stimulate participant thinking to set up reflection, and eliciting discussion to promote change talk among group participants. The facilitator also provided a space for educators to reflect on their professional values, discuss specific topics (e.g., addressing the access gap), share ideas, and problem-solve barriers to adopting new practices. The BASIS-T facilitator used group-based motivational interviewing (MI; Miller & Rollnick, 2012) techniques (i.e., open-ended questioning, reflective listening, ruler questions, pros/cons; Magill & Hallgren, 2019) to elicit change-talk and promote collective self-efficacy.

In the post-training session, which occurred immediately after the final day of EBPP training (i.e., GBG in this study)

and lasted around 60 min. During the tailored post-session, participants were first asked to respond to a single-item dichotomous measure capturing their intentions to implement the EBPP. Depending on their answer, teachers convened into two smaller groups. Those who indicated they had intentions to implement the GBG began implementation planning, and those who indicated they did not have intentions to implement or are ambivalent engaged in a further motivational experience that involved normalizing ambivalence via ruler questions and a decisional balance activity to explore pros and cons of changing or not by trying out small components of the EBPP.

Teachers also received a digitally delivered booster 15 days after the beginning of the school year (self-paced, e-course experience developed in Articulate 360) that provided a tailored experienced based on (a) whether they had initiated EBPP delivery or not, (b) whether they had intentions to begin using GBG, and (c), whether they had no intention of using GBG.

The BASIS-T trainer (Author 2) was a doctoral level licensed school psychologist, trained in school-based MI (Frey et al., 2017), with extensive expertise and years of experience supporting educators to prevent and address student SEB problems. Moreover, the project also included a consultant who is a trainer with the Motivational Interviewing Network of Trainers (MINT; motivationalinterviewing.org) and has a history of conducting MI research (Hartzler et al., 2007).

Active Comparison Control Group

Teachers randomly assigned to the AC group received a 3-h pre-training session prior to GBG training, designed to control for dose and delivery of information. The AC facilitator defined, described, and advocated for EBPP implementation in schools and used an educational approach that emphasized didactic delivery of content with opportunities for teachers to reflect on the information that was shared. Teachers in the AC condition also participated in a 1-h post-training, which involved them reviewing and discussing the importance of EBP implementation and reviewing the definition and dimensions of fidelity. Finally, teachers received an e-booster 15 days after the start of the school year that was not tailored and revisited core concepts of the GBG. The trainer for the AC condition was carefully selected as a comparable match to the BASIS-T trainer in both skill in delivery of training content, demographics (e.g., age, race, gender), qualifications (PhD level) and training style.

Table 1 Participant demographics

	Total		Control		BASIS-T	
	N	%	N	%	N	%
Total	81	100	39	100.0	42	100.0
Primary race						
AI/AN	2	2.3	0	0.0	2	4.8
Asian	1	1.1	0	0.0	1	2.4
Latinx	2	2.3	0	0.0	2	4.8
White	83	94.3	39	100.0	37	88.1
Secondary race (multiracial)						
White	2	66.7	0	0.0	2	4.8
Other	1	33.3	1	2.6	0	0.0
Gender						
F	82	93.2	37	94.9	39	92.9
M	6	6.8	2	5.1	3	7.1
Highest ed						
Bachelors	28	16.4	11	28.2	15	35.7
Masters	60	68.2	28	71.8	27	64.3
Grade taught						
K & 1st	34	44.2	14	35.9	20	47.6
2nd & 3rd	22	35.1	10	25.6	12	28.6
4th & 5th	17	10.4	12	30.8	5	11.9
Other (SPED, Art, music, reading)	8	3.9	3	7.7	5	11.9
	M	(SD)	M	(SD)	M	(SD)
Years teaching	16.0	(15.5)	15.8	(15.2)	15.3	(15.8)
Baseline assessments						
Attitudes towards EBPs	4.37	(0.43)	4.43	(0.37)	4.33	(0.47)
Outcome expectancy	5.92	(0.86)	5.90	(0.86)	5.93	(0.88)
Ownership/role	3.49	(0.95)	3.76	(.0.94)	3.24	(0.89)
Subjective social norms	5.21	(0.67)	5.22	(0.61)	5.19	(0.73)
General self-efficacy	5.74	(0.85)	5.72	(0.84)	5.75	(0.86)
Task self-efficacy	6.02	(0.82)	6.10	(0.78)	5.95	(0.86)
Intentions to implement	6.56	(0.52)	6.51	(0.59)	6.60	(0.46)

Note. AI = American Indian, AN = Alaskan Native

Procedures

Recruitment and Randomization

Institutional Review Board approval was obtained by the university human subjects committee and the school district research department. Recruitment procedures began through communications with district leadership regarding the needs within elementary schools and GBG. This led to conversations with elementary principals regarding the nature of the project and providing opportunities to ask questions. Interested principals, met with their teaching staff to identify teachers who indicated an interest in receiving free training in an EBPP. Participating schools were randomly assigned to the BASIS-T condition or the AC condition after being paired via a Euclidean distance nearest neighbor matching analysis using variables related to enrollment size, percent of students receiving free and reduced priced meals, and percent of students of color. Each school had a single best matching school with one exception, which was assigned to its second and third closest matching schools, as they were a match, to facilitate a three-way match with the smallest overall Euclidean distance. Within these pairs and the triple match, we randomly assigned to one of the two conditions.

Post-assignment, there were no statistically significant differences on any of the matching variables, no differences on several other student variables (i.e., percent of English language learners, qualified for special education, homeless), and no differences on percent of teachers with an advanced degree. Teachers were contacted via email to obtain consent and a link to the pre-training survey via Qualtrics. Online pre- and post-training surveys were collected from the AC condition and BASIS-T condition. Participants had one week prior to the pre-training session and one week after post-training session to complete surveys. Eighty percent of participants completed post-surveys within 2 days of the post-training session. Teachers received $140 for participating in training and $50 for each wave of data collection.

Condition Delivery Procedures

Teachers in both conditions received GBG training and pre- and post-training experiences at separate locations on the same day. The GBG training was provided by a third-party purveyor of the GBG. Teachers were combined by condition across schools for the 2-day training. Teachers in the BASIS-T condition ($n = 42$) were exposed to the following sequence of training: Day 1: BASIS-T pre-training session, GBG training part 1; Day 2: GBG training part 2, BASIS-T post-training session. Teachers in the AC condition ($n = 39$) were exposed to the same sequence as the BASIS-T group, only the pre- and post-training sessions involved AC activities.

Measurement Procedures

Demographics were collected at pre-training. Full-scale mechanism data was collected via the Research Electronic Data Capture (REDCap) at pre- and post-training (August, 2019). To capture the impact of BASIS-T over time, teachers also received REDCap surveys each month following post-training between September and February 2020, and one final administration in April, 2020. Data collection was interrupted due to COVID-related school closures in March, 2020. For each separate section throughout the survey, teachers received specific prompts that provided guidance for responding (e.g., "*Please select the statement that most closely reflects your implementation of the good behavior game*"; "*Think back over this past month. On average, how many days during a school week did you play the actual game part of the good behavior game?*").

Measures

A detailed description of study measures, including reliabilities in the current sample, is provided in Supplemental File 2. We selected measures that allowed us to capture various mechanisms of behavior change that align with behavior

change theory (i.e., attitudes toward EBP, ownership/role, outcome expectancies, self-efficacy, social norms, intentions to implement; Ajzen, 1985, Ajzen & Manstead, 2007; Schwarzer et al., 2011), implementation outcomes (i.e., adoption, degree of implementation, GBG fidelity, BASIS-T procedural reliability), and student behavioral outcomes. Regarding mechanisms, we selected existing measures with strong psychometric properties. For mechanisms where no measures existed, we constructed them based on guidelines for developing psychometrically sound measures of TPB constructs (Francis et al., 2004). The survey items for adoption, fidelity, and degree of implementation were developed based on prominent implementation outcome and fidelity literature as well as with existing questionnaires (Carroll et al., 2007; Perepletchikova & Kazdin, 2005, Proctor et al., 2011; Sanetti & Kratochwill, 2009). Finally, student behavioral outcomes were measured via teacher-reported class-wide Direct Behavior Rating (DBR; Sims et al., 2021). To reduce participant response burden, a subset of 2–6 items for each mechanism scale were administered at the monthly surveys (Fricker et al., 2014). Items that were selected had the highest shared variance with all items in the scale based on their squared multiple correlations at pre- and post-training, and all reliabilities among reduced subscales correlated 0.89 or above with full scales. The measurement reliabilities, correlations between full and partial item measures, and the time-points at which each were captured are displayed in Table 2.

Mechanisms

To measure attitudes toward evidence-based practice, eleven items from the school-adapted version of the Evidence-Based Practice Attitudes Scale (EBPAS; Cook et al., 2018). Items were endorsed on a 5-point Likert-type scale ranging from "Not at All" to "Very great Extent" (subscale α range = 0.75–0.91).

The outcome expectancies measure assessed the degree to which teachers believed evidence-based classroom management practices would result in positive outcomes (e.g., "*Evidence-based classroom management practices offer significant potential to improve outcomes for students*"). Outcome Expectancies were endorsed on a 7-point Likert-type scale ranging from "Completely Disagree" to "Completely Agree" (α = 0.89–0.94).

The ownership or role in managing behavior and teaching appropriate behavior in the classroom measure (e.g., "*Student behavior is the responsibility of parents—not teachers*"). This construct was measured on a 7-point Likert-type scale ranging from "Completely Disagree" to "Completely Agree" (α = 0.71–0.82).

The modified subjective norms measure is an 8-item measure used to capture two types of EBPP implementation-related subjective norms: injunctive (what a social group

would approve of) and descriptive (how a social group actually behaves). Injunctive norms are the perception of what ought to be or what the social group would approve of (e.g., *"Others who I respect expect me to adopt and implement evidence-based practices that promote students' SEB functioning"*). Injunctive norm items (4 items, α = 0.71–0.90) are rated on a seven-point scale (-3 = "I should", 0 = "Neutral", 3 = "I Should Not"). Descriptive norms describe perceptions of how the social group actually does things (e.g., *"Practitioners like me find the time to implement evidence-based practices"*). Descriptive norm items (3 items, α = 0.69–0.84) were rated on a seven-point scale (-3 = "Strongly agree", 0 = "Neutral", 3 = "Strongly disagree").

Two measures of perceived behavioral control (i.e., self-efficacy) were used. The General self-efficacy measure (α = 0.92). asked about teachers' general confidence in using evidence-based practices (e.g., *"When I try really hard, I am able to overcome barriers to implement evidence-based practices that help me manage classroom behavior problems"*). The Task self-efficacy measure (α = 0.89–0.90) captures tasks associated with positive classroom behavior management generally (e.g., *"I am confident I can routinely provide neutral redirection, correction, and/or performance-specific feedback to increase my students' awareness of and engagement in classroom rules and expectations"*). Items are rated on a 7-point Likert-type scale ranging from "Completely Disagree" to "Completely Agree."

The Intentions to Use measure was used to capture teachers' intentions to implement specific practices associated with positive classroom behavior management (e.g., *"I have every intention of implementing the following practice: Positive reinforcement system is set up which allows students to access rewarding experiences when following classroom rules and expectations regarding behavior"*). The scale includes 9 items (four general items, α = 0.98; five GBG-specific items; α = 0.74–0.88) and is rated on a 7-point Likert-type scale ranging from "Completely Disagree" to "Completely Agree".

Implementation Outcomes

Adoption, fidelity, and degree of implementation were also measured via REDCap. We measured whether the teacher began using the GBG (*adoption;* 1 = yes, 0 = no), and whether the intervention components or practices were delivered as intended and with the correct dose and frequency (*fidelity*). Teachers who endorsed implementing the GBG completed a self-rated 14-item measure about the GBG practices they performed. Items were on a 3-point scale (1 = Done consistently and appropriately, 2 = Done inconsistently, 3 = Not done). Some items were more general to good classroom management (e.g., *"Identify privileges, activities, and rewards to be earned"*), and some items were more GBG-specific (e.g., *"Announce and*

Table 2 Measure reliabilities, correlations, and time of collection

Measure	Item Reliability (*a*)	Correlation between full and partial item measure (*r*)	Full-Scale Time of Measurement	Partial-Scale Time of Measurement
EBPAS	–	0.95	Pre- and Post-Training, Post-study (M8)	M1–M6
Openness	0.85–0.89	–	–	–
Appealing	0.75–0.85	–	–	–
Fit	0.88–00.91	–	–	–
Outcome expectancies	0.89–0.94	1.00	Pre- and Post-Training, Post-study (M8)	M1–M6
Descriptive norms	0.69–0.84	0.89	Pre- and Post-Training, Post-study (M8)	M1–M6
Injunctive norms	0.71–0.90	0.90	Pre- and Post-Training, Post-study (M8)	M1–M6
General self-efficacy	0.92	0.95	Pre- and Post-Training, Post-study (M8)	M1–M6
Task self-efficacy	0.89–0.90	0.94	Pre- and Post-Training, Post-study (M8)	M1–M6
Ownership/Role	0.71–0.82	0.91	Pre- and Post-Training, Post-study (M8)	M1–M6
Specific intention	0.74–0.88	0.91	Post-Training, Post-study (M8)	M1–M6
Degree of implementation	–	–	M1–M8	–
Adoption	–	–	M1–M8	–
General fidelity	0.62–0.64	–	M2, M4, M6, M8	–
GBG fidelity	0.76.91	–	M2, M4, M6, M8	–
BASIS-T fidelity	0.88	–	After Post-Training	–
DBR	0.79	–	M8	–

EBPAS = Evidence-Based Practice Attitudes Scale; GBG = The good behavior game; M = Month following post-training; DBR = Direct Behavior Rating; all correlations between full and partial items *p* < .001

celebrate winning teams only"). An exploratory factor analysis using maximum likelihood extraction and rotating factor loadings using Varimax with Kaiser normalization identified two factors, with items loading in perfect alignment with those hypothesized: General fidelity (6 items) and GBG-specific fidelity (8 items). Therefore, we selected the GBG-specific fidelity items to analyze and report on for this manuscript. Teachers who did not implement at that timepoint had every GBG-specific item replaced with a 3 ("not done").

Degree of implementation was measured every other month, with teachers responding on a 5-point scale, with lower scores indicating greater likelihood of implementation: 1 = Currently implementing at least some parts of GBG, 2 = Started implementing GBG but stopped entirely, 3 = Have not started implementing GBG but intend to, 4 = Have not started implementing GBG and uncertain if want to, 5 = Have not started implementing GBG and do not intend to.

Student Behavioral Outcomes

We used DBR to measure student outcomes. Teachers were provided with operational definitions and examples of the three target behaviors in question (i.e., on-task, prosocial, and disruptive behaviors). They were asked to reflect on a typical day in their classroom and rate the proportion of time that students were on-task and engaging in prosocial behaviors as well as the number of disruptive behaviors that their students engaged in (Sims et al., 2021).

BASIS-T Procedural Fidelity Assessment

Delivery of BASIS-T was recorded and Authors 1 and 4 independently rated procedural fidelity to capture the degree to which the core features of BASIS-T were delivered as planned (adherence to content, group-based motivational interviewing, and treatment differentiation). Results indicated that the three BASIS-T components were delivered as planned and are reported more extensively in Larson et al. (2021).

Data Analytic Approach

Demographics and descriptive variables were analyzed using means, standard deviations, and group proportions. Cross-tabulations with chi-square tests were used to test condition differences in proportion to who adopted the GBG immediately post-training and who adopted GBG at any point. T-tests compared conditions on student behavior outcomes. Spaghetti plots were created for each mechanism and implementation outcome variable to visually assess the participant and condition time trend trajectories (Swihart et al., 2010).

Mixed effects modeling was the primary method of testing longitudinal research questions (Magezi, 2015). We applied random effect terms to adjust standard errors due to nesting by timepoint within teacher within school for longitudinal variables. For each dependent variable (implementation status, GBG fidelity, frequency) an initial model was computed that included covariates for condition and time, followed by models that added the interaction of condition and time, quadratic time, and an interaction of quadratic time by condition (Nagle, 2018). In most models, time was centered at pre-test. Time was centered at the first post-training monthly survey for implementation status, GBG fidelity, and delivery amounts, because that was the first time that participants were surveyed about implementation. Model fit was used as the criteria for research question significance testing (Nagle, 2018). We only report here on the best fitting models, selected via the chi-square, Akaike Information Criteria, and Bayesian Information Criteria, with significance and smaller values indicating better model fit (Nagle, 2018). Adoption was analyzed as a two-level mixed effects model, without time, using a Bernoulli distribution link function. Due to the relatively small sample size, restricted maximum likelihood estimation was applied to obtain model coefficients; as appropriate full maximum likelihood was used to obtain model fit statistics. Nesting was found within the implementation status variable (ICC = 0.19). All other intra-class correlations among mechanism and fidelity variables were less than 0.05 with nonsignificant deviance tests, and therefore random effects were fixed in these models to permit convergence.

For the exploratory tests of the basic elements of mediation analyses, we computed Pearson's correlations and point-biserial correlations between task self-efficacy, implementation status immediately post-training, implementation status at any time during follow-up, student on-task behavior, student prosocial behavior, and student disruptive behavior (van Kestern & Oberski, 2019). Due to the pilot, exploratory nature of the study, we did not adjust for familywise error rate (Feise, 2002).

Effect Sizes

Because pilot studies are underpowered by definition, our study aims to avoid Type II false negatives in preparation for a larger trial. Therefore, we report unstandardized effect sizes and interpret marginal findings in the results below (Lee et al., 2014). We caution readers against drawing strong inference. Reporting effect sizes is crucial for interpretating applied research results and for determining practical significance within intervention science (Kelley & Preacher, 2012). However, clear guidelines for reporting effect size in multilevel models have not been provided (Peng et al., 2013). Non-standardized effect size estimates

are produced with multilevel models. While non-standardized effect sizes make it difficult to compare magnitude of effect across outcomes, they do provide information about the directionality of the effect. Thus, in this study, for both non-significant and significant findings, we report on the directionality of the beta estimates provided through multilevel modeling (Table 3). We report Cohen's *d* effect sizes for student behavior outcome data (Cohen, 1988).

Power Analyses

For our primary research question (RQ 1), this pilot trial is powered to detect a minimum detectable standardized effect size of 0.73 to 0.76. This assumes stratified block randomization at the school level, grade-level cluster Intra-Class Correlation ranging from 0 to 0.15, nine schools with nine teachers per school, and school-level effect size variability of 0.01. As is the norm for pilot studies (Leon et al., 2011), this study is underpowered when considering the complexity of the full analytic model and the multiple paths/tests. To be powered to detect a more moderate effect ($d = 0.5$), the trial would need to include approximately 15 schools with nine teachers each.

Missing Data

A missing data analysis was completed to determine whether survey completion was balanced between conditions. We calculated the proportion of participants who completed each monthly survey by condition. We also sought to determine whether differences at baseline mechanism score correlated with whether participants responded to surveys. Missing data was deleted listwise after selecting mechanisms of interest in order to retain as much data as possible.

Results

Baseline mean and standard deviation scores for all mechanism variables are displayed in Table 1. The conditions were highly similar on all variables. Model coefficients are displayed in Table 3. Results from the missing data analysis indicated that there was equal missing data between conditions. On average, participants in both conditions completed 72% of monthly surveys. Across the five monthly surveys, 75% of participants completed each monthly survey on average (range: 64–87%). Correlations between survey completion and mechanism score indicated weak correlations ($M = 0.010$, range $= -0.0095–0.141$). Due to the nature of this pilot study being exploratory, we focus our results on both statistically and marginally significant findings.

RQ1. Overall, model findings suggested that there was one significant effect (task self-efficacy) and one marginal effect (attitudes toward EBP) between the two conditions

Table 3 Longitudinal Mixed Effects Models Predicting Change Over Time by Condition

Outcome	Intercept	Time	BASIS-T	BASIS-T × Time	Time^2	BASIS-T × Time^2
Mechanisms						
Attitudes towards EBPs	4.344***	− 0.067***	− 0.003	0.032	–	–
Outcome expectancies	5.882***	− 0.084**	0.229	0.019	–	–
Ownership/role	3.797***	0.025	− 0.472*	0.037	–	–
Subjective social norms	5.313***	− 0.063**	0.063	0.024	–	–
General self-efficacy	5.707***	− 0.338***	0.137	0.065	–	–
Task self-efficacy	6.009***	− 0.307**	0.088	0.289*	0.037*	− 0.055*
Intentions to implement	6.405***	− 0.168***	0.135	0.012	–	–
Implementation Variables						
Implementation status	2.589***	− 0.439***	− 0.812*	0.134*	0.114**	–
GBG Fidelity	1.621***	0.162*	0.161	− 0.191	–	–
Delivery times per day	3.807***	0.079	0.192	− 0.079	–	–
Delivery times per week	5.706***	− 0.106	1.197	− 0.196	–	–

* $p < .05$, ** $p < .01$, *** $p < .001$

Note. Mechanism variables were time-centered at pre-training timepoint. Implementation variables were time-centered at the first monthly survey, which was the first timepoint at which they were administered. Quadratic time trends are displayed if deviance statistics (χ^2, Akaike, and Bayesian Information Criterion) indicated a significantly better model fit than linear trend. All intercepts are centered at the first timepoint, with the exception of implementation status, which is set at the first timepoint post-training at which participants could respond. EBPAS = Evidence Based Practices Attitudes Scale; GBG = Good Behavior Game
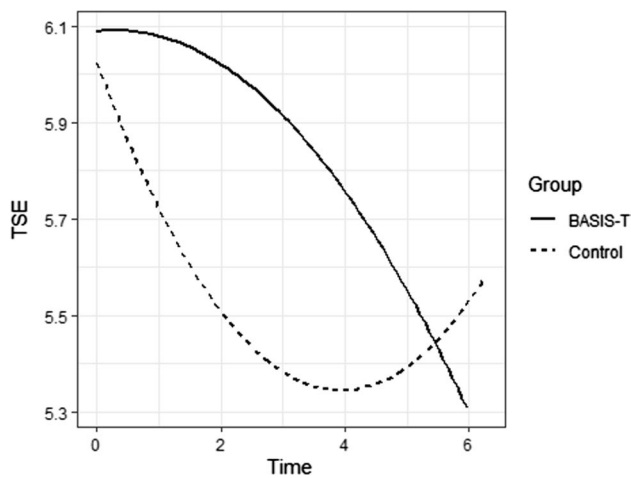
Fig. 2 Task self-efficacy estimated change over time by month. *Note* Time point '0' indicates immediate post-training score

Fig. 3 Degree of implementation estimated change over time by month. *Note* Lower scores indicate greater implementation

on dependent variables representing mechanisms of behavioral change. Task self-efficacy increased significantly from pre- to post-training for teachers in the BASIS-T condition while teachers in the AC condition deteriorated during this timeframe (Time Est = − 0.307, $p = 0.003$; Time × Condition Est = 0.289, $p = 0.033$; Time^2 Est = 0.036, $p = 0.047$; Time^2 x Condition Est = − 0.055, $p = 0.025$; Model Chi square = 5.32, $p = 0.069$) (Fig. 2). Task self-efficacy was the only behavior change mechanism variable that had a significant quadratic time trend, therefore results below discuss linear time trends. General self-efficacy scores decreased over time for both conditions (Est = − 0.338, $p < 0.001$) and there was no difference between conditions across time. Total scores on the EBPAS significantly deteriorated over time (Est = − 0.067, $p < 0.001$), and differences by condition over time approached significance in favor of BASIS-T (Est = 0.032, $p = 0.084$). Both groups had outcome expectancy scores that significantly deteriorated over time (Est = − 0.084, $p = 0.003$) and there were no differences by condition. The BASIS-T condition had a significantly lower ownership/role score on the pre-test (Est = − 0.472, $p = 0.028$), and there were no differences in change over time overall or by condition. Both groups had subjective norms scores that significantly deteriorated over time (Est = − 0.063, $p = 0.009$), and there were no differences by condition over time. When examining effect size estimates for each of the DVs representing mechanisms of behavior change, results indicated that all estimates were in the hypothesized direction favoring the BASIS-T condition. On average, both conditions demonstrated a reduction in specific intentions score over time (Est = − 0.168, $p < 0.001$). There were no differences between conditions at post-training or on rate of change over time on intentions to implement.
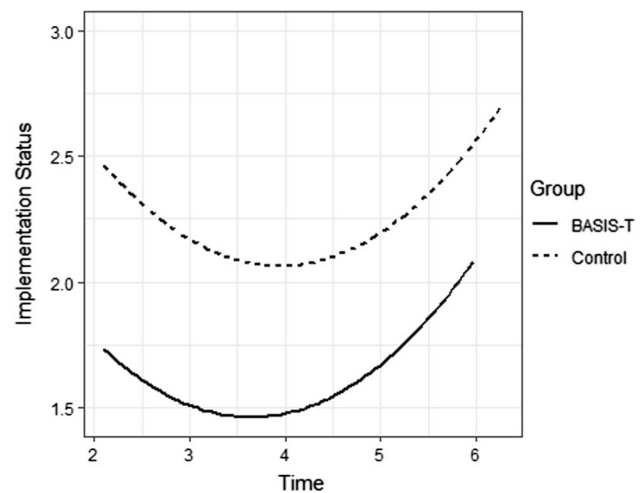
RQ2. Our analyses found an immediate effect of BASIS-T on self-reported adoption of GBG. Crosstabulations with chi-square tests found a significant difference between conditions on the percentage of participants who adopted the GBG immediately post-training (AC $n = 12$, 40.0%; BASIS-T $n = 28$, 73.7%; $\chi^2(1) = 7.853$, $p = 0.005$). However, by the end of the academic year there were minimal differences between the groups in the percentage of participants who *ever* adopted any of the GBG practices (AC $n = 26$, 74.3%; BASIS-T $n = 33$, 82.5%; $\chi^2 = 0.751$, $p = 0.386$). This was consistent with mixed effects modeling, which found that degree of implementation was significantly better for the BASIS-T condition immediately after training (Est = − 0.812, $p = 0.012$; see Fig. 3). Both conditions' degree of implementation improved over time (Time Est = − 0.438, $p < 0.001$), and a significant and positive quadratic timepoint indicated flattening over time for both groups (Est = 0.114, $p = 0.003$). On average, the BASIS-T group had less improvement over time (from follow-up 1 to 5) than the AC group (Est = 0.134, $p = 0.043$) given the higher proportion of early adopters immediate post training.

There were no differences in self-reported intervention fidelity for GBG-specific practices immediately post training, and both conditions' fidelity improved over time (Est = 0.162, $p = 0.029$). There was a marginal trend in favor of BASIS-T such that intervention fidelity remained steady for BASIS-T while worsening for the AC condition (Est = − 0.191, $p = 0.057$). For teachers delivering the GBG, there were no differences between conditions in the number of times the game was played per day at the first timepoint post-training no changes over time, and no differences between conditions. There was a borderline finding in favor of the BASIS-T condition that indicated this condition may have played the game about one day per week more often

Table 4 Correlations among hypothesized and significant mechanisms and student outcomes

	Task self-efficacy post-training	Implemented GBG immediately post-training	Ever implemented GBG	Student on-task behavior	Student prosocial behavior	Student disruptive behavior
Task self-efficacy post-training	–	0.222a	0.213a	0.283*	0.263*	− 0.079
Implemented GBG immediately post-training	–	–	0.581**	0.259*	0.207a	0.006
Ever implemented GBG	–	–	–	− 0.037	0.164	0.163

Tests used Pearson correlations for associations between continuous variables (task self-efficacy and student behavior) and Pearson point-biserial correlations for associations between continuous and dichotomous variables (implementation status). $a = p < .10$

than the AC condition at the first timepoint post-training ($Est = 1.197$, $p = 0.097$). There were no other differences in the number of times per week the game was played.

RQ3. Conditions did not statistically differ on DBR scores, though effect sizes were in the favor of the BASIS-T condition, for all subscales including student on-task behaviors (Cohen's $d = − 0.368$ [95% CI: − 0.836, 0.102], $t_{(70)} = -1.55$, $p = 0.125$), student prosocial behaviors (Cohen's $d = − 0.274$ [95% CI: − 0.740, 0.194], $t_{(70)} = − 1.16$, $p = 0.252$), and student disruptive behaviors (Cohen's $d = 0.380$ [, $t_{(70)} = 1.60$, $p = 0.114$). Again, these findings should be interpreted with caution given they were not significant at the $a = 0.05$ level.

RQ4. Table 4 displays exploratory analyses of the associations among the mechanism, implementation, and student outcome variables. Exploratory analyses indicated a possible small association between task self-efficacy immediately post-training and whether the GBG was implemented immediately post-training ($r = 0.222$, $p = 0.069$), and whether the GBG was ever implemented ($r = 0.213$, $p = 0.081$). There were statistically significant associations between task self-efficacy and student on-task behavior ($r = 0.283$, $p = 0.022$), and between task self-efficacy and student prosocial behavior ($r = 0.263$, $p = 0.034$), but no association between task self-efficacy and student disruptive behavior ($r = − 0.079$, $p = 0.529$). Implementing the GBG immediately post training was associated with ever implementing the GBG ($r = 0.581$, $p < 0.001$), student on-task behavior ($r = 0.259$, $p = 0.037$), possibly associated with student prosocial behavior ($r = 0.207$, $p = 0.098$), but not associated with student disruptive behavior ($r = 0.006$, $p = 0.962$). Having ever implemented the GBG was not associated with student on-task behavior ($r = − 0.037$, $p = 0.758$), student pro-social behavior ($r = 0.164$, $p = 0.169$), or student disruptive behavior ($r = 0.163$, $p = 0.173$).

Discussion

The uptake and high-fidelity delivery of universal EBPPs enhances the public health impact of prevention science in schools. Cornerstone didactic training is a key implementation strategy used by schools to facilitate EBPP uptake and use (Joyce & Showers, 2002; Owens et al., 2014); however, due to the disappointing outcomes it tends to yield, it was our task to develop a theory-based strategy to boost the effects of training on teacher uptake and use of universal EBPPs and analyze the underlying mechanisms of action to understand how and why the effects occurred. This study used a hybrid type III, cluster randomized pilot trial design to examine the effects of BASIS-T on teacher implementation outcomes and observe student classroom behavior. Findings revealed that those who received BASIS-T were associated with a 34% increase in immediate adoption of a class-wide EBP compared to the comparison group. We also discuss the nuanced effects of BASIS-T on behavioral mechanisms and provide directions for future research and practice.

Two principles guided our interpretation of the findings. First, we grounded our interpretation in the fact that this was a pilot trial. Pilot trials are crucial for translational research processes because they detect potential signals that could inform planning of a larger, more rigorous, and adequately powered study (Eldridge et al., 2016). Consequently, we discuss both significant and marginal findings below as they elucidate potential signals of effects to inform research questions, measurement approach, and analytic models for a larger trial. Second, BASIS-T was evaluated in the context of a universal intervention. In a review of multiple meta-analyses of universal prevention programs, Tanner-Smith and colleagues (2018) found that, across universal prevention programs targeting school-aged youth, median average effect sizes fell between Cohen's $d = 0.07$ and 0.16, with smaller effects on behavior. They concluded that the statistical benchmarks outlined in by Cohen (1988) for small (0.20), medium (0.50), and large (0.80) effects are inappropriate to use in interpreting findings from universal interventions.

Thus, effects on student outcomes were interpreted in context of the estimates from Tanner-Smith et al. (2018).

Results from the current study contribute to a nuanced understanding of BASIS-T effects. First, BASIS-T teachers had significantly higher task self-efficacy and marginally higher attitudes towards EBPPs post-training than AC teachers. Task self-efficacy differs from general self-efficacy as it relates to confidence in one's ability to perform specific behaviors (e.g., specific practices that constitute the GBG) rather than general confidence. These findings align with other research indicating that attitudes and self-efficacy impact behavior change (Sheeran et al., 2016). Second, teachers in the BASIS-T condition reported significantly greater early adoption compared to teachers in the AC condition following training. For example, 74% of teachers in the BASIS-T condition indicated they were early adopters of the GBG practices compared to only 40% in the AC condition. Across both conditions, exploratory analyses indicated that teachers with higher task self-efficacy were more likely to be early adopters, which correlated positively with student on task behaviors and negatively with student disruptive behaviors (Table 4). Conversely, having ever adopted was not associated with student outcomes. Third, we found that teachers' ratings deteriorated across all mechanisms of behavior change over the academic year, suggesting that as time goes on, teachers are likely to develop less favorable attitudes, lower social norms, and weaker self-efficacy related to EBPP implementation. Although the BASIS-T group was initially associated with significant differences in outcome expectancies (Larson et al., 2021), these deteriorated over time. Fourth, regarding self-reported intervention fidelity, we found a marginal trend in favor of BASIS-T, indicating that fidelity remained relatively steady for BASIS-T, while weakening for the AC condition. There was a marginal finding in favor of the BASIS-T condition indicating that teachers in this condition reported that they played the game about one day per week more than the AC condition post-training. Last, student behavioral outcomes (on task, prosocial, disruptive) favored the BASIS-T condition, and had effect sizes of $d = 0.27$–0.38, which fall above the median average effect range for universal prevention programs on various outcomes. These findings, although not statistically significant, support the logical argument that earlier adoption to evidence-based behavioral interventions corresponds with improved behavioral outcomes. However, given the nonsignificant value in this study, additional research directly or indirectly associating higher fidelity with student outcomes is needed.

Implications for BASIS Theory of Change

This is the third study analyzing the effects of the BASIS strategy, which enables us to interpret our findings in the context of the results from previous studies. It appears that BASIS has an effect on initial adoption of EBPPs as well as on specific proximal mechanisms of behavior change, including task self-efficacy ($p < 0.05$) and potentially on outcome expectancies and attitudes toward EBPs ($p = 0.08$) (This study and Cook et al., 2015; Larson et al., 2021). In turn, increased task self-efficacy appears to increase the early yield of training on practitioner adoption, as significantly more practitioners who receive BASIS initiate implementation right away. However, in response to consultation, practitioners in the AC group are likely to catch up with those in the treatment group. While the proportion of practitioners adopting an EBPP become more similar to that of the BASIS-T condition over the course of the year, the sooner in the year practitioners adopt the EBPP, the greater the likelihood of improved student outcomes. Teachers' attitudes toward an EBP are also highly dependent on the specific practices within the selected EBP as well as the process of selection, i.e., were teachers viewpoints considered when choosing the EBP? Favorable attitudes have been shown to impact teachers' decisions to adopt a given practice (Sheeran et al., 2016).

BASIS may have a small, or even negligible, effect on sustained intervention fidelity, although adequately powered studies with rigorous observation of fidelity are needed. BASIS is designed to act on antecedent mechanisms of behavior change by engaging practitioners in a group-based motivational experience prior to and immediately after EBPP training. Fidelity, on the other hand, involves behavioral persistence after initial adoption has occurred, which is exemplified by practitioners' effort to continuously improve their delivery of an EBPP with fidelity. Fidelity is largely the focus of implementation strategies like consultation, fidelity audits, and protected time for reflection and planning. The findings from the BASIS studies suggest that researchers need to design more precise implementation strategies that are tailored and designed to effect changes in specific implementation outcomes of interest throughout the implementation process. For example, strategies designed to promote adoption (i.e., initiation or starting to implement) may function differently than those aiming to enhance sustained fidelity. Implementation strategies are infrequently designed with such precision in mind, potentially attenuating the observed effects of some implementation efforts (Lewis et al., 2018). To address this, future iterations of BASIS will attempt to study and impact the extent to which teachers engage in existing, ongoing supports to promote sustained behavior change.

The findings from this study suggest the need for multifaceted implementation approaches that combine and sequence different strategies to promote successful adoption and persistence towards high fidelity. The implementation literature is mixed with regard to whether multi-faceted or

bundled strategies are more effective than single component, discrete strategies (Merle et al., 2022a, 2022b; Squires et al., 2014). Regardless, we concur with others who have articulated how the effects of multi-faceted strategies can be increased through precisely targeting mechanisms of behavior change (Lewis et al., 2018), and through careful sequencing of strategies at critical junctures of the implementation process (Kilbourne et al., 2014). For example, engaging practitioners in motivational experiences prior to training may prime them to be more engaged and responsive to training. In turn, volitional implementation strategies situated immediately post training, such as implementation planning, are well-positioned to help motivated practitioners initiate the adoption of EBPPs. Next, post-training consultation can be deployed once practitioners have had a chance to begin implementing an EBPP to support them to persist towards high fidelity, and make fidelity-consistent adaptations to fit the context.

Implications for Research and Practice

BASIS-T serves as a novel implementation strategy targeting motivational processes linked with individual behavior change in a group-based context that is designed to complement training and consultation. This study elucidates the importance of attending to task self-efficacy as a mechanism of behavior change related to adoption for prevention programming. This study necessitates a closer look at determinants of sustainability, i.e., what happens over the course of a school year or beyond that leads to the deterioration in behavior change mechanisms among implementers. Post-training implementation strategies should attend to this phenomenon and buffer implementers against experiences that weaken factors related to behavior change. Coaching grounded in motivational interviewing offers a useful approach to sustain motivational factors to behavior change once implementers experience the daily demands of their jobs (Frey et al., 2020).

Limitations and Future Directions

There are limitations of this study that warrant discussion and pinpoint directions for future research. First, this was a pilot trial with a relatively small sample of schools and teachers. Therefore, interpretations of marginally significant predictors were explored to inform future iterations of BASIS as well as the theory of change and should be interpreted with caution. Further, randomization occurred at the school level, which limits statistical power. This decision, however, was made to increase feasibility and reduce the risk of contamination because one mechanism of BASIS is social norms. Although we explored the association between

significant mechanisms (i.e., task self-efficacy) and adoption, we had limited power to perform more robust mediation and moderation analyses. Future studies may consider teacher-level randomization with contamination prevention measures.

Another limitation to this work was the reliance on self-reported collection of implementation outcome data. Self-reported fidelity data alone tends to be overestimated compared to permanent product review and direct observation (McKenna et al., 2014). Therefore, future research testing the BASIS strategy should incorporate a more robust metric of implementation outcomes, such as observational coding of implementer behavior or permanent product review.

BASIS-T had an effect on immediate adoption and revealed potential mechanisms impacting this change. However, to ensure the sustainability of change efforts, future strategies that uncover the mechanisms that protect against implementer fade and regression toward the mean are warranted, as there is scant research in this area (Merle et al., 2022b). In this study, we attempted to protect against implementer deterioration in fidelity by including a timely booster experience roughly two weeks into the school year, which aligns with other research documenting this phenomenon (Noell et al., 2005). Future research should directly target determinants of successful implementation across the implementation process to ensure maintenance of effects. For instance, future iterations of BASIS intend to capture the degree to which teachers engage in ongoing supports, such as coaching and consultation. These data were not collected in this trial, which is a limitation.

Research findings from the various iterations of BASIS continually reveal limited variability in intentions to implement at the individual level, which may be due to social desirability or the way in which behavioral intention items are constructed that results in participants endorsing responses on the higher end of the scale. Developing measures to capture the full range of variability between participants in their status on behavior change is critical to explain variance in these constructs as well as predict variability in implementation outcomes (e.g., adoption and fidelity). Implementation science is still nascent, and researchers need to continually focus on establishing a common set of instruments to capture behavior change constructs related to the adoption and delivery of EBPPs that are validated for school contexts. One avenue explored by Moullin and colleagues (2018), involves incorporating the Rasch measurement theory when developing innovation-specific intentions, which offers advantages to classical factor analysis, such as being able to assess the appropriateness of response options.

Our implementation strategy focused on individual-level mechanisms of behavior change. Results of the pilot study provide generalizable knowledge that self-efficacy is an important mechanism for initiating a new behavior. Indeed,

a variety of inner and outer-setting factors interact to determine the sustained use of innovations in schools (Owens et al., 2014). The fit of the innovation, the implementation climate, perceptions of support from colleagues and leaders, external policy and pressures to fulfill other obligations, and myriad other factors are known to contribute to the successful implementation of a new practice that must be identified and targeted in tandem with individual-level strategies (Flottorp et al., 2013). Typically, best practice in implementation science includes a carefully planned, dynamic collaborative process between stakeholders and implementation experts who assist in the selection of an innovative practice based on need (Wolfenden et al., 2021), and the utilization of a variety of strategies targeting the various socioecological levels of an organization at specific time points to achieve implementation success (Aarons et al., 2011; Fixsen et al., 2005). BASIS-T was developed to increase the yield of professional development in schools as a pre-implementation strategy. Since single implementation strategies likely only achieve a slice of the whole effect, future studies and real-world implementation efforts need to look at the efficacy of BASIS within a multi-level, blended approach during active implementation and sustainment phases to capitalize on the early gains and achieve implementation success.

There is enough preliminary evidence supporting BASIS that there is now a need for a larger efficacy trial enabling analysis of with whom, under what conditions, and how or why BASIS works. For instance, it is unknown whether participation in BASIS increases engagement in other evidence-based implementation strategies (e.g., performance feedback), which are important determinants for sustained fidelity (Stormont et al., 2015). Future research on BASIS should look at more proximal process outcomes, such as whether BASIS leads to greater engagement in the EBP training provided by the purveyor group as well as embedded consultation within the school setting. This study was also significantly limited due to interruptions caused by the COVID-19 pandemic. This halted the EBPP purveyor from collecting observational fidelity data for the later data points during the school year, and ultimately limited the amount of data we were able to collect and analyze. Findings related to self-reported fidelity may not hold for "gold standard" observed fidelity (Sanetti & Kratochwill, 2009, 2011).

Conclusion

Successful implementation of EBPPs boils down to behavior change, and motivation is a critical antecedent mechanism of behavior change. Among the behavior change constructs, we believe findings from this study and other studies suggest that task self-efficacy in particular serves as an important

antecedent mechanism of behavior change that drives initial adoption. Our work on BASIS has continually been in pursuit of a mechanistic understanding of what, how, and why change occurs for whom and under which circumstances. We urge researchers to carefully target and monitor early adoption and intervention fidelity and consider the specific strategies that operate on each of these important implementation outcomes. This study's findings advance understanding of the type of implementation strategies that complement pre-implementation training and post-training consultation in schools. Indeed, no single strategy guarantees implementation success, however, we encourage future research that continues to develop and test implementation strategies that clearly specify their change mechanisms at various levels and time points during implementation.

Declarations

Conflict of interest The authors have no conflicts of interest to declare. Institutional Review Board approval was obtained by the university human subjects committee and the school district research department, and all procedures complied with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Prior informed consent was obtained for all participants involved.

References

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

Ajzen, I, & Manstead, A. (2007). Changing health-related behaviors: An approach based on the theory of planned behaviour the scope of social psychology: Theory and applications. In *The Scope of Social Psychology: Theory and Applications* (pp. 43–63). Psychology Press.

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In *Action control: From cognition to behavior* (pp. 11–39). Springer Berlin Heidelberg.

Barrish, H. H., Saunders, M., & Wolf, M. M. (1969). Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis, 2*(2), 119–124. https://doi.org/10.1901/jaba.1969.2-119

Beidas, R. S., Edmunds, J. M., Marcus, S. C., & Kendall, P. C. (2012). Training and consultation to promote implementation of an empirically supported treatment: A randomized trial. *Psychiatric Services, 63*(7), 660–665. https://doi.org/10.1176/appi.ps.201100401

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: Science and Practice, 17*(1), 1–30. https://doi.org/10.1111/j.1468-2850.2009.01187.x

Bowman-Perrott, L., Burke, M. D., Zaini, S., Zhang, N., & Vannest, K. (2016). Promoting positive behavior using the good behavior game: A meta-analysis of single-case research. *Journal of Positive Behavior Interventions, 18*(3), 180–190. https://doi.org/10.1177/1098300715592355

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2*(1), 40. https://doi.org/10.1186/1748-5908-2-40

Cohen, J. (1988). The effect size. In *statistical power analysis for the behavioral sciences* (pp. 77–83).

Collier-Meek, M. A., Sanetti, L. M. H., & Fallon, L. M. (2017). Incorporating applied behavior analysis to assess and support educators' treatment integrity. *Psychology in the Schools, 54*(4), 446–460. https://doi.org/10.1002/pits.22001

Cook, C. R., Davis, C., Brown, E. C., Locke, J., Ehrhart, M. G., Aarons, G. A., Larson, M., & Lyon, A. R. (2018). Confirmatory factor analysis of the Evidence-Based Practice Attitudes Scale with school-based behavioral health consultants. *Implementation Science, 13*(1), 116. https://doi.org/10.1186/s13012-018-0804-z

Cook, C. R., Lyon, A. R., Kubergovic, D., Browning Wright, D., & Zhang, Y. (2015). A supportive beliefs intervention to facilitate the implementation of evidence-based practices within a multitiered system of supports. *School Mental Health, 7*(1), 49–60. https://doi.org/10.1007/s12310-014-9139-3

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

Domitrovich, C. E., Bradshaw, C. P., Greenberg, M. T., Embry, D., Poduska, J. M., & Ialongo, N. S. (2010). Integrated models of school-based prevention: Logic and theory. *Psychology in the Schools, 47*(1), 71–88. https://doi.org/10.1002/pits.20452

Eldridge, S. M., Chan, C. L., Campbell, M. J., Bond, C. M., Hopewell, S., Thabane, L., & Lancaster, G. A. (2016). CONSORT 2010 statement: Extension to randomised pilot and feasibility trials. *BMJ*. https://doi.org/10.1136/bmj.i5239

Fallon, L. M., Collier-Meek, M. A., Kurtz, K. D., & DeFouw, E. R. (2018). Emailed implementation supports to promote treatment integrity: Comparing the effectiveness and acceptability of prompts and performance feedback. *Journal of School Psychology, 68*, 113–128. https://doi.org/10.1016/j.jsp.2018.03.001

Fallon, L. M., Collier-Meek, M. A., Maggin, D. M., Sanetti, L. M., & Johnson, A. H. (2015). Is performance feedback for educators an evidence-based practice? A systematic review and evaluation based on single-case research. *Exceptional Children, 81*(2), 227–246.

Feise, R. J. (2002). Do multiple outcome measures require p-value adjustment? *BMC Medical Research Methodology, 2*, 8. https://doi.org/10.1186/1471-2288-2-8

Fixsen, D. L., Naoom, S.F., Blase, K. A., Friedman, R. M., Wallace, F., Burns, B., ... & Shern, D. (2005). Implementation research: Asynthesis of the literature. https://nirn.fpg.unc.edu/sites/nirn.fpg.unc.edu/files/resources/NIRN-MonographFull-01-2005.pdf

Flottorp, S. A., Oxman, A. D., Krause, J., Musila, N. R., Wensing, M., Godycki-Cwirko, M., Baker, R., & Eccles, M. P. (2013). A checklist for identifying determinants of practice: a systematic review and synthesis of frameworks and taxonomies of factors that prevent or enable improvements in healthcare professional practice. *Implementation Science, 8*(1), 1–11.

Francis, J., Eccles, M. P., Johnston, M., Walker, A., Grimshaw, J. M., Foy, R., Kaner, E. F., Smith, L., & Bonetti, D. (2004). *Constructing questionnaires based on the theory of planned behaviour: A manual for health services researchers*. Centre for Health Services Research, University of Newcastle. http://openaccess.city.ac.uk/1735/

Frey, A. J., Lee, J., Small, J. W., Sibley, M., Owens, J. S., Skidmore, B., Johnson, L., Bradshaw, C. P., & Moyers, T. B. (2020). Mechanisms of motivational interviewing: A conceptual framework to guide practice and research. *Prevention Science*. https://doi.org/10.1007/s11121-020-01139-x

Frey, A. J., Lee, J., Small, J. W., Walker, H. M., & Seeley, J. R. (2017). Motivational interviewing training andassessment system (MITAS) for school-based applications. *Report on Emotional & Behavioral Disorders in Youth, 17*(4), 86–92.

Fricker, S., Yan, T., & Tsai, S. (2014). Responseburden: What predicts it and who is burned out. In JSM proceedings (pp. 4568–4577). http://www.asasrms.org/Proceedings/y2014/files/400298_500838.pdf

Hagger, M. S., & Chatzisarantis, N. L. D. (2014). An integrated behavior change model for physical activity. *Exercise and Sport Sciences Reviews, 42*(2), 62–69. https://doi.org/10.1249/JES.0000000000000008

Hartzler, B., Baer, J. S., Dunn, C., Rosengren, D. B., & Wells, E. (2007). What is seen through the looking glass: The impact of training on practitioner self-rating of motivational interviewing skills. *Behavioural and Cognitive Psychotherapy, 35*(4), 431–445. https://doi.org/10.1017/S1352465807003712

Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development* (Vol. 3). Alexandria, VA: Association for Supervision and Curriculum Development.

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods, 17*(2), 137–152. https://doi.org/10.1037/a0028086

Kilbourne, A. M., Almirall, D., Eisenberg, D., Waxmonsky, J., Goodrich, D., Forney, J., Kirchner, J., Solberg, L., Main, D., Bauer, M., Kyle, J., Murphy, S., Nord, K., & Thomas, M. (2014). Protocol: Adaptive implementation of effective programs trial (ADEPT): Cluster randomized SMART trial comparing a standard versus enhanced implementation strategy to improve outcomes of a mood disorders program. *Implementation Science., 9*, 132. https://doi.org/10.1186/s13012-014-0132-x

Landes, S. J., McBain, S. A., & Curran, G. M. (2019). An introduction to effectiveness-implementation hybrid designs. *Psychiatry Research, 280*, 112513. https://doi.org/10.1016/j.psychres.2019.112513

Larson, M., 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. (2021). Examining the effects of a brief, group-based motivational implementation strategy on mechanisms of teacher behavior change. *Prevention Science*. https://doi.org/10.1007/s11121-020-01191-7

Lee, E. C., Whitehead, A. L., Jacques, R. M., & Julious, S. A. (2014). The statistical interpretation of pilot trials: Should significance thresholds be reconsidered? *BMC Medical Research Methodology, 14*(1), 41. https://doi.org/10.1186/1471-2288-14-41

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research, 45*(5), 626–629. https://doi.org/10.1016/j.jpsychires.2010.10.008

Lewis, C. C., Klasnja, P., Powell, B. J., Lyon, A. R., Tuzzio, L., Jones, S., Walsh-Bailey, C., & Weiner, B. (2018). From classification to causality: Advancing understanding of mechanisms of change in implementation science. *Frontiers in Public Health*. https://doi.org/10.3389/fpubh.2018.00136

Lyon, A. R., & Bruns, E. J. (2019). From evidence to impact: Joining our best school mental health practices with our best implementation strategies. *School Mental Health, 11*(1), 106–114. https://doi.org/10.1007/s12310-018-09306-w

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 of an evidence-based trauma intervention. *Implementation Science, 14*(1), 54. https://doi.org/10.1186/s13012-019-0905-3

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." *Administration and Policy in Mental Health and Mental Health Services Research, 44*(1), 16–28. https://doi.org/10.1007/s10488-015-0650-0

Lyon, A. R., Whitaker, K., Locke, J., Cook, C. R., King, K. M., Duong, M., Davis, C., Weist, M. D., Ehrhart, M. G., & Aarons, G. A. (2018). The impact of inter-organizational alignment (IOA) on implementation outcomes: Evaluating unique and shared organizational influences in education sector mental health. *Implementation Science, 13*(1), 24. https://doi.org/10.1186/s13012-018-0721-1z

Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: An introductory tutorial and free, graphical user interface (LMMgui) [Mini Review]. *Frontiers in Psychology.* https://doi.org/10.3389/fpsyg.2015.00002

Magill, M., & Hallgren, K. A. (2019). Mechanisms of behavior change in motivational interviewing: Do we understand how MI works? *Current Opinion in Psychology, 30*, 1–5.

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

Merle, J. L., Larson, M. F., Cook, C. R., Brewer, S. K., Hamlin, C., Duong, M., McGinnis, J. L., Thayer, A. J., Gaias, L. M., & Lyon, A. R. (2022a). A mixed-method study examining solutions to common barriers to teachers' adoption of evidence-based classroom practices. *Psychology in the Schools.* https://doi.org/10.1002/pits.22732

Merle, J. L., Thayer, A. J., Larson, M. F., Pauling, S., Cook, C. R., Rios, J. A., McGinnis, J. L., & Sullivan, M. M. (2022b). 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.

Miller, W. R., & Rollnick, S. (2012). *Motivational interviewing: Helping people change.* Guilford press.

Moullin, J. C., Ehrhart, M. G., & Aarons, G. A. (2018). The role of leadership in organizational implementation and sustainment in service agencies. *Research on Social Work Practice, 28*(5), 558–567.

Mouzakitis, A., Codding, R. S., & Tryon, G. (2015). The effects of self-monitoring and performance feedback on the treatment integrity of behavior intervention plan implementation and generalization. *Journal of Positive Behavior Interventions, 17*(4), 223–234. https://doi.org/10.1177/1098300715573629

Nadeem, E., Saldana, L., Chapman, J., & Schaper, H. (2018). A mixed methods study of the stages of implementation for an evidence-based trauma intervention in schools. *Behavior Therapy, 49*(4), 509–524. https://doi.org/10.1016/j.beth.2017.12.004

Nagle, C. (2018, September). An Introduction to fitting and evaluating mixed-effects models in R. In *Annual Pronunciation in Second Language Learning and Teaching Conference* (p. 82).

Noell, G. H., Gansle, K. A., Mevers, 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.

Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L., 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. https://doi.org/10.1080/02796015.2005.12086277

Owens, J. S., Lyon, A. R., Brandt, N. E., Masia Warner, C., 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.

Peng, C.-Y.J., Chen, L.-T., Chiang, H.-M., & Chiang, Y.-C. (2013). The impact of APA and AERA guidelines on effect size reporting. *Educational Psychology Review, 25*(2), 157–209. https://doi.org/10.1007/s10648-013-9218-2

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice, 12*(4), 365–383.

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

Proctor, E., Silmere, H., Raghavan, 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.

Proctor, E. K., Powell, B. J., & McMillen, J. C. (2013). Implementation strategies: Recommendations for specifying and reporting. *Implementation Science, 8*(1), 139. https://doi.org/10.1186/1748-5908-8-139

Reinke, W. M., Herman, K. C., & Sprick, R. (2011). *Motivational interviewing for effective classroom management: The classroom check-up* (Riley-Tillman TC, ed.). Guilford press.

Reinke, W. M., Lewis-Palmer, T., & 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

Sanetti, L. M. H., & Collier-Meek, M. A. (2019). *Supporting successful interventions in schools: Tools to plan, evaluate, and sustain effective implementation.* Guilford Publications.

Sanetti, L., Fallon, L., & Collier-Meek, M. A. (2013). Increasing teacher treatment integrity through performance feedback provided by school personne. *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 treatment 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. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38*(4), 445.

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

Schwarzer, R., Lippke, S., & Luszczynska, A. (2011). Mechanisms of health behavior change in persons with chronic illness or disability: The Health Action Process Approach (HAPA). *Rehabilitation Psychology, 56*(3), 161–170. https://doi.org/10.1037/a0024509

Sheeran, P., Maki, A., Montanaro, E., Avishai-Yitshak, A., Bryan, A., Klein, W. M., Miles, E., & Rothman, A. J. (2016). The impact of changing attitudes, norms, and self-efficacy on health-related intentions and behavior: A meta-analysis. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association, 35*(11), 1178–1188. https://doi.org/10.1037/hea0000387

Sims, W. A., King, K. R., Reinke, W. M., Herman, K., & Riley-Tillman, T. C. (2021). Development and preliminary validity evidence for the direct behavior rating-classroom management (DBR-CM). *Journal of Educational and Psychological Consultation, 31*(2), 215–245. https://doi.org/10.1080/10474412.2020.1732990

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, 4*, 160–175.

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 health-care professionals' behaviours? An overview of systematic reviews. *Implementation Science, 9*(1), 1–22.

Stirman, S. W., Kimberly, J., Cook, N., Calloway, A., Castro, F., & Charns, M. (2012). The sustainability of new programs and innovations: A review of the empirical literature and recommendations for future research. *Implementation Science, 7*(1), 1–19. https://doi.org/10.1186/1748-5908-7-17

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

Swihart, B. J., Caffo, B., James, B. D., Strand, M., Schwartz, B. S., & Punjabi, N. M. (2010). Lasagna plots: A saucy alternative to spaghetti plots. *Epidemiology (cambridge, Mass), 21*(5), 621–625. https://doi.org/10.1097/EDE.0b013e3181e5b06a

Tanner-Smith, E. E., Durlak, J. A., & Marx, R. A. (2018). Empirically based mean effect size distributions for universal prevention programs targeting school-aged youth: A review of meta-analyses. *Prevention Science, 19*(8), 1091–1101. https://doi.org/10.1007/s11121-018-0942-1

van Kesteren, E., & Oberski, D. (2019). Exploratory mediation analysis with many potential mediators. *Structural Equation Modeling: A Multidisciplinary Journal, 26*(5), 710–723. https://doi.org/10.1080/10705511.2019.1588124

Wolfenden, L., Foy, R., Presseau, J., Grimshaw, J. M., Ivers, N. M., Powell, B. J., Taljaard, M., Wiggers, J., Sutherland, R., Nathan, N., Williams, C. M., Kingsland, M., Milat, A., Hodder, R. K., & Yoong, S. L. (2021). Designing and undertaking randomised implementation trials: Guide for researchers. *BMJ.* https://doi.org/10.1136/bmj.m3721