# A pilot study comparing tools for tracking implementation strategies and treatment adaptations

Callie Walsh-Bailey[1,2] iD, Lorella G Palazzo[1],
Salene MW Jones[3], Kayne D Mettert[1] iD, Byron J Powell[2] iD,
Shannon Wiltsey Stirman[4,5] iD, Aaron R Lyon[6] iD,
Paul Rohde[7] and Cara C Lewis[1,6] iD

## Abstract

**Background:** Tailoring implementation strategies and adapting treatments to better fit the local context may improve their effectiveness. However, there is a dearth of valid, reliable, pragmatic measures that allow for the prospective tracking of strategies and adaptations according to reporting recommendations. This study describes the development and pilot testing of three tools to be designed to serve this purpose.

**Methods:** Measure development was informed by two systematic reviews of the literature (implementation strategies and treatment adaptation). The three resulting tools vary with respect to the degree of structure (brainstorming log = *low*, activity log = *moderate*, detailed tracking log = *high*). To prospectively track treatment adaptations and implementation strategies, three stakeholder groups (treatment developer, implementation practitioners, and mental health providers) were randomly assigned one tool per week through an anonymous web-based survey for 12 weeks and incentivized to participate. Three established implementation outcome measures, the Acceptability of Intervention Measure, Intervention Appropriateness Measure, and Feasibility of Intervention Measure, were used to assess the tools. Semi-structured interviews were conducted to gather more nuanced information from stakeholders regarding their perceptions of the tools and the tracking process.

**Results:** The three tracking tools demonstrated moderate to good acceptability, appropriateness, and feasibility; the activity log was deemed the most feasible of the three tools. Implementation practitioners rated the tools the highest of the three stakeholder groups. The tools took an average of 15 min or less to complete.

**Conclusion:** This study sought to fill methodological gaps that prevent stakeholders and researchers from discerning which strategies are most important to deploy for promoting implementation and sustainment of evidence-based practices. These tools would allow researchers and practitioners to track whether activities were treatment adaptations or implementation strategies and what barrier(s) each targets. These tools could inform prospective tailoring of implementation strategies and treatment adaptations, which would promote scale out and spread.

[1]Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA
[2]Brown School, Washington University in St. Louis, St. Louis, MO, USA
[3]Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[4]National Center for PTSD, VA Palo Alto Health Care System, Menlo Park, CA, USA
[5]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA

[6]Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, WA, USA
[7]Oregon Research Institute, Eugene, OR, USA

**Corresponding author:**
Callie Walsh-Bailey, Brown School, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130, USA.
Email: callie.w@wustl.edu

**Plain Language Summary**
Strategies to support the implementation of evidence-based practices may be more successful if they are carefully customized based on local factors. Evidence-based practices themselves may be thoughtfully changed to better meet the needs of the settings and recipients. This study reports on a pilot study that aimed to create various types of tools to help individuals involved in implementation efforts track the actions they take to modify and implement interventions. These tools allow individuals to track the types of activities they are involved in, when the activities occurred, who was involved in the implementation efforts, and the reasons or rationale for the actions. The three tools in this study used a combination of open-ended and forced-response questions to test how the type of data recorded changed. Participants generally found the tools quick and easy to use and helpful in planning the delivery of an evidence-based practice. Most participants wanted more training in implementation science terminology and how to complete the tracking tools. Participating mental health providers would have liked more opportunities to review the data collected from the tools with their supervisors to use the data to improve the delivery of the evidence-based practice. These tools can help researchers, providers, and staff involved in implementation efforts to better understand what actions are needed to improve implementation success. Future research should address gaps identified in this study, such as the need to involve more participants in the tool development process.

## Background

Mounting evidence suggests that tailoring implementation strategies, which involves assessing relevant determinants and designing the strategy to address key determinants and stakeholder needs in a given context, may be more effective than standardized approaches (Baker et al., 2015; Kilbourne et al., 2014; Kirchner et al., 2017; Lewis et al., 2015; Powell et al., 2017). Furthermore, treatments may be optimized when they are adapted, that is, when thoughtful, deliberate modifications are made to the components of a treatment or how it is delivered, with the goal of enhancing treatment fit or effectiveness (Baumann et al., 2017; Chambers et al., 2013; Wiltsey Stirman et al., 2017, 2019). Together, tailoring implementation strategies and adapting treatments may improve the sustainability of evidence-based practices (EBPs; Chambers et al., 2013). Tailoring and adapting are often reactive processes, occurring in an unplanned manner throughout implementation (Wensing, 2017), but may best achieve their desired effect through thoughtful planning (Baumann et al., 2017; Chambers et al., 2013). Both tailoring strategies and adapting treatment are complex processes that may benefit from tools to prompt and guide the consideration of implicated factors, but few such tools exist.

Moreover, careful measurement of these tailoring and adapting processes is needed to evaluate their independent and interactive impact on targets and outcomes of interest. Currently, there are no valid, reliable, pragmatic tools that allow for planning or tracking strategy tailoring and treatment adaptations simultaneously. Two studies offer preliminary methods for tracking implementation strategies (Boyd et al., 2017; Bunger et al., 2017). First, Bunger and colleagues (2017) created an activity log in which participants in implementation efforts recorded implementation activities, their purpose, actors involved, and time spent on the activity. This log intended to collect retrospective and prospective data in a low-burden manner. Data extracted from this tool allowed for the identification of the action, actor, temporality, and dose of the implementation strategy. However, reporting according to recommendations (Proctor et al., 2013) is limited, as the justification, strategy target, and outcome are not explicit. These early efforts confirm the importance of contemporaneous tracking given problems with retrospective recall (Bunger et al., 2017).

Second, Boyd et al. (2017) developed a coding system to track implementation strategies and map them onto reporting recommendations (Proctor et al., 2013) that captured the same categories as the activity log (Bunger et al., 2017), as well as the strategy target, justification, and outcome. In theory, the coding system generates sufficient detail to allow the implementation strategy to be reproduced, but this must be tested. Practical limitations imposed by coding, such as expertise needed and feasibility outside the research context, should be addressed through future work.

Although adapting treatment is considered as one of the more than 73 implementation strategies (Powell et al., 2015), a review of treatment adaptations (Wiltsey Stirman et al., 2013) and resulting Framework for Reporting Adaptations and Modifications-Expanded (FRAME; Wiltsey Stirman et al., 2019) illustrates a host of complexities that would not be captured by existing tracking methods. To understand how adaptations influence clinical and implementation outcomes, a feasible and reliable tracking tool that explicitly focuses on treatment adaptation may be useful. Without tracking, we are unable to discern if an adaptation constitutes drift or leads to EBP improvements (Chambers & Norton, 2016).

There are calls in implementation science for more detailed reporting of both implementation strategies (Rudd et al., 2020; Wilson et al., 2017) and treatment adaptations (Baumann et al., 2017; Wiltsey Stirman et al., 2013). Persistent knowledge and methodological gaps limit the degree to which researchers and practitioners can prospectively inform implementation tailoring and treatment adaptations. Without tools for prospective tracking, the field is hampered in its ability to disentangle which activity is most important, to what degree, and when for promoting sustainment. The perspectives of implementation practitioners, clinicians, and others involved in implementation efforts are crucial to informing, evaluating, replicating, and scaling these efforts. Thus, any tool to support planning or tracking implementation strategies and treatment adaptations in real time should engage stakeholders (Glasgow & Riley, 2013); however, existing methods have largely been researcher driven. One such tool, developed by Rabin and colleagues (2018), utilizes FRAME to track adaptations made to treatments and implementation strategies in real time. As these authors note, the divisions between implementation strategies and treatments are at times murky; Eldh et al. (2017) suggest greater clarity between these two would benefit research and practice alike. Distinguishing between implementation strategies and treatments, and changes made to each, is necessary to determine whether observed outcomes are attributable to treatments or the strategies used to implement them (Fixsen et al., 2013; National Academies of Sciences Engineering and Medicine; Division of Behavioral and Social Sciences and Education; Board on Children, Youth, and Families; Committee on Fostering Healthy Mental, Emotional, and Behavioral Development Among Children and Youth, 2019).

The long-term objective of our work is to generate valid, reliable, and pragmatic tools that can be utilized in a variety of contexts by diverse end-users to prospectively guide and contemporaneously track implementation strategy tailoring and treatment adaptations. We report on a mixed-methods pilot test of three tools to prospectively track implementation strategies and treatment adaptations. The aims of the pilot study were to (1) provide an initial assessment of the utility, acceptability, feasibility, and appropriateness of each tool, (2) assess the extent to which stakeholders adapted the EBP of interest and the nature of those adaptations, and (3) gather preliminary evidence of the validity and reliability of the three tools. The results of this pilot will inform subsequent refinement of the tools for optimal impact.

## Method

### Tool development

The study's principal investigator (PI) and first author reviewed available literature to locate examples of tracking tools. We modified two existing tools (Boyd et al., 2017; Bunger et al., 2017) and developed a third, less structured tool based on examples from journaling prompts and brainstorming activities (Albright & Cobb, 1988; Tuckett & Stewart, 2004). The varying degree to which tools were open-ended or structured was intended to test potential trade-offs related to the degree of structure used (i.e., open versus forced). The continuum of structure may impact the ease of response and participant burden, as well as completeness, quality, and actionability of the data collected. Greater reliability is associated with structured tools due to the standardization of response options (Edwards, 2010). Less structured tools allow for greater content validity and comprehensive qualitative coverage of a construct because participants are not limited to options considered by the developer (Demetriou et al., 2015). Although open-ended questions allow for more diverse collection of responses, these require more extensive coding and are more prone to missing data (Reja et al., 2003). Open-ended questions may increase participant burden, while close-ended questions may be subject to bias, whether imposed by the investigator's lens or participant avoidance of extreme options (Edwards, 2010).

Despite differing structures, all three tools sought to capture the same categories of information (Table 1). The tools were informed and evaluated by three implementation frameworks and compilations (Supplemental File 1). First, Proctor and colleagues (2013) offered recommendations for implementation strategy reporting and specification to enable replication (Figure 1). Our tools varied in the degree to which responses aligned with these recommendations. Second, the School Implementation Strategies, Translating ERIC Resources (SISTER) compilation revised implementation strategy labels and definitions to improve fit with school settings (Cook et al., 2019). Only the most structured tool offered standardized strategy labels; the SISTER compilation was applied to evaluate response quality from the other two tools. Third, FRAME provided a system for classifying and reporting treatment adaptations and included the identification of who made the modification, at what level of delivery modifications are made, whether the modification was made to the content, context, or training and evaluation of the intervention, and the nature of the context or content modifications (Wiltsey Stirman et al., 2019). Similar to the application of the SISTER compilation, treatment adaptation categories were embedded in the most structured tool and used to evaluate responses to the other two tools. Tool content and design were developed through an iterative process informed by research team meetings with implementation practitioners and the EBP developer. The piloted tools described below are available in Supplemental Files 2–4.

1. Brainstorming log: This tool was the most open-ended (Table 1). The brainstorming log was

informed by a vocational education trainer's log (Albright & Cobb, 1988) and literature on journaling as a qualitative data collection method (Tuckett & Stewart, 2004). It consisted of six questions, one of which was multiple choice ("what is your role") and five free text. Before describing their activities, participants indicated the range of dates for which they reported. First, participants reported treatment adaptations made, describing content and context modifications in separate questions. Second, participants reported on barriers encountered and strategies deployed (or proposed) to address those barriers.

2. Activity log: This tool, based on Bunger et al. (2017), was moderately structured and open-ended (Table 1), using five same questions: participant role, date of the activity, time spent on the activity, the purpose, and the attendees. Unique to this tool, we asked about the intended outcome of the activity. This tool did not require participants to specify their activity as an implementation strategy or treatment adaptation.

3. Detailed tracking log: This tool was the most structured and detailed (Table 1). In addition to the questions from the activity log, participants categorized each activity as an implementation strategy or treatment adaptation through pre-populated response options. Implementation strategies were organized into nine categories delineated by Waltz et al. (2015) and assigned a label from the SISTER compilation (Cook et al., 2019). Participants reported on treatment adaptations according to FRAME (Wiltsey Stirman et al., 2019).

## Measures

Three measures of implementation outcomes, the Acceptability of Intervention Measure (AIM), Intervention Appropriateness Measure (IAM), and Feasibility of Intervention Measure (FIM; Weiner et al., 2017), were used to assess the likelihood that stakeholders might adopt these tools. Each contains four items rated on a 5-point scale (1 = *completely disagree*, 5 = *completely agree*). Summary scores for each measure were created by averaging responses, with higher values reflecting more favorable perceptions. AIM measured the degree to which each tracking tool was satisfactory to stakeholders (Cronbach's $\alpha = .97$). IAM measured the relevance or perceived fit of each tool (Cronbach's $\alpha = .97$). FIM assessed the degree to which each tool could be successfully utilized (Cronbach's $\alpha = .96$). These scales were followed by an open-ended question, "Please tell us why you rated this tracking method the way you did. What did you like/not like about it?"

The 6-item Adaptations to Evidence-Based Practices Scale (AES) explored treatment adaptations using an established quantitative measure (Lau et al., 2017) as a concurrent validity assessment. This measure includes six items rated on a 5-point scale (0 = *not at all*, 4 = *a very great extent*). The AES contains two subscales, "augmenting" adaptations ("I modify how I present or discuss components of the EBP") and "reducing/reordering" ("I shorten/condense pacing of the EBP"). Mean scores were calculated for each AES subscale, with higher scores indicating more adaptation (overall Cronbach's $\alpha = .94$, augmenting subscale = .90, reducing/reordering subscale = .93).

## Pilot testing

*Setting and participants.* Our study capitalized on the implementation of the Blues Program, an evidence-based cognitive behavioral group depression indicated prevention program. This EBP intends to promote engagement in pleasant activities and reduce negative cognitions among teens at risk of developing major depression (Stice et al., 2008, 2010). A non-profit that offers services to children and families included an Implementation Support Center that provided oversight of the Blues Program implementation in New York state high schools. The implementation practitioner team at the Implementation Support Center developed an implementation plan which included predetermined strategies to guide EBP implementation across all participating schools; school-based providers delivering the intervention came up with additional ad hoc implementation strategies. The Blues Program trained school-based mental health providers to facilitate group sessions, and the developer consulted with implementation practitioners during program initiation. These three mutually exclusive stakeholder groups, the Blues Program developer, implementation practitioners, and mental health providers, participated in our pilot by reporting on their Blues Program-related activities through the tracking tools. The Blues Program developer ($N = 1$) was a PhD trained investigator with 30 years of post-training experience. The implementation practitioners ($N = 3$) all held master's degrees and had an average of 5.5 years of professional experience. The school-based mental health providers ($N = 7$) also all had master's level training and worked in their profession an average of 4.3 years. This study was deemed exempt by the Institutional Review Board; informed consent was not obtained.

*Data collection.* The tools were administered to participants across two 6-week cycles of the Blues Program. All groups were randomly assigned one tool per week, distributed through an email link to a web-based survey so that each tool was administered twice each cycle. Participants were instructed to complete tracking by reflecting on the

prior week's activities; they were given 6 days to record activities. Each survey included the IAM, FIM, AIM, and AES. A US$10 per survey incentive was offered during the first round of data collection, which increased to US$20 for the second data collection cycle. Response rates are reported in Table 2.

After tracking data collection concluded, the first author conducted semi-structured interviews with participants: treatment developer ($N = 1$), implementation practitioners ($N = 2$), and school-based mental health providers ($N = 5$). One implementation practitioner was on leave when interviews were conducted and two providers left their positions and could not be contacted. Response rate among remaining participants was 100% ($N = 8$). Interviews allowed for in-depth exploration into stakeholders' experience with the tools. A semi-structured interview guide (Supplemental File 5) was prepared to capture information on (1) perceived benefits of tracking, (2) tracking method preferences, (3) tracking process, (4) background/training for tracking, (5) tracking execution and completion, (6) general utility of tracking, and (7) contextual information. Participants were emailed each tool for reference during the interview. Participants received US$40 on interview completion. Interviews were recorded and transcribed.

### Data analysis

To assess participant perceptions of the tools, we compared scores on the AIM, IAM, and FIM using a generalized estimating equation (GEE, a type of multilevel model). Analyses were completed at the response level, rather than at the individual participant level. Scores on the measures were nested within weeks and roles, and we examined the fixed effects for role (treatment developer, implementation practitioners, and mental health providers) and tool (detailed tracking log, activity log, and brainstorming log). As the measures did not show substantial skew or kurtosis, we used a linear model that assumes a normal distribution. We also used a similar analysis to see if the AES differed by role and tool. Because each of the predictors of interest had three categories, we ran each GEE twice, changing the reference group so we could examine all pairwise comparisons (i.e., activity log with detailed tracking log, detailed with brainstorming log, and brainstorming log with activity log). For the analyses, scores were only used if participants had reported at least 50% of the measure items. Four records were excluded because of missing data for a final sample of 59 responses.

We entered tool responses into an Access database for coding and created a codebook using established implementation frameworks and compilations. We categorized implementation strategies using the ERIC categories (Waltz et al., 2015) and SISTER compilation (Cook et al., 2019), reporting recommendations for strategy specification (Proctor et al., 2013) and FRAME (Wiltsey Stirman et al., 2019) for treatment adaptations. Two research specialists conducted dual independent coding and met weekly to resolve discrepancies. If consensus was not reached, the PI made a final decision. Total coding and consensus time varied by number of activities reported, ranging from 10 to 25 min per response. Coding times did not vary significantly across tools. These data offered a characterization of the reported activities based on alignment with implementation strategy reporting recommendations (Proctor et al., 2013) and FRAME (Wiltsey Stirman et al., 2013) (Table 3).

The first and second authors conducted dual independent coding of all interview transcripts using ATLAS.ti (Version 7.1; Ringmayr, 2013). A codebook containing a priori codes was developed and iteratively refined with emergent codes. Coders held weekly consensus meetings to resolve discrepancies and reach agreement on emergent codes (Hill et al., 2005). Code reports were obtained and analyzed for main themes and illustrative quotes.

## Results

### Acceptability, appropriateness, and feasibility

Participants rated the tracking methods as having moderate to good acceptability, appropriateness, and feasibility (Figure 2). The activity log was significantly more acceptable ($b = -.407$, $p = .006$) and feasible ($b = -.517$, $p < .001$) but not more appropriate ($b = -.174$, $p = .394$) than the detailed tracking log. The brainstorming log did not significantly differ on acceptability, appropriateness, and feasibility from the other two methods (all $p$s $> .09$), except participants perceived the activity log as more feasible than the brainstorming log ($b = -.346$, $p = .015$; see Figure 2).

As shown in Figure 2, mental health providers tended to report lower acceptability, appropriateness, and feasibility of all tracking methods compared to treatment developer and implementation practitioners (all $p$s $< .05$ except the comparison of treatment developer with providers on appropriateness, $p = .122$). However, the means on the AIM, IAM, and FIM for mental health providers were between 3 and 4 (on a 1–5 scale), indicating that they generally approved of the tracking methods.

### Adaptation to EBP

The extent to which participants reported making adaptations to the Blues Program was low (Figure 3). The nature of adaptations was more frequently augmenting factors rather than reducing/reordering. Responses to the AES in the activity log indicated the least amount of adaptation. This is consistent with the coding of the activity log data, which revealed only one treatment adaptation across 17 responses, compared to 4 in the detailed tracking log and 20 in the brainstorming
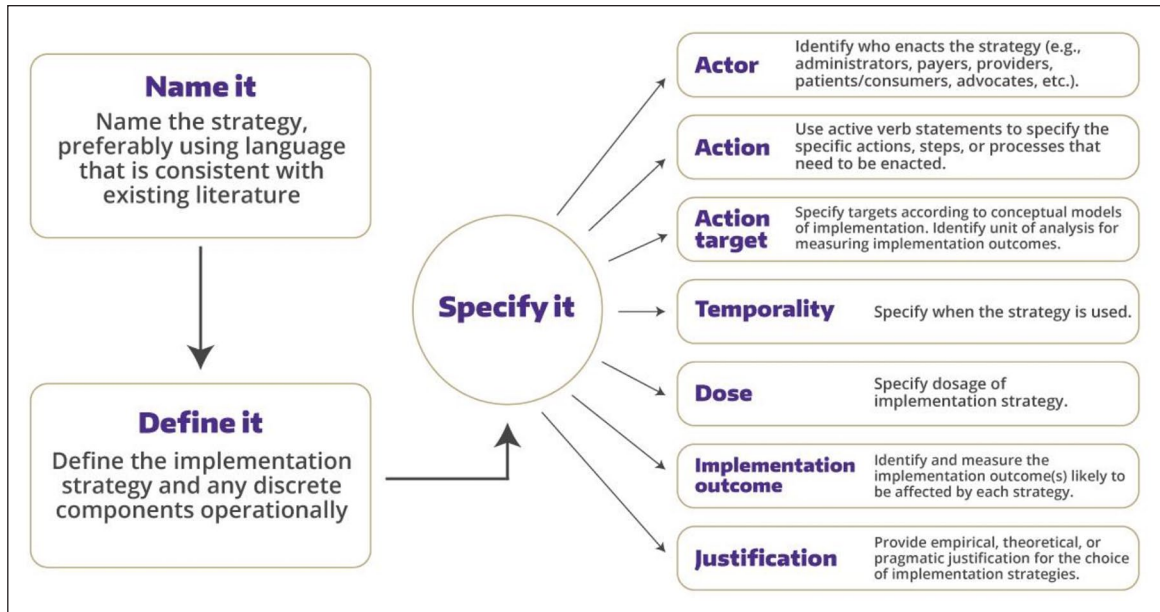
**Figure 1.** Implementation strategy reporting recommendations.
Source: Adapted from Proctor et al. (2013). Reproduced with Permission from Bryan Weiner, PhD.

**Table 1.** Structure of self-report measures and reporting recommendations by tracking tool.

| | | Not structured ⟶ Highly structured | | | | | |
|---|---|---|---|---|---|---|---|
| | | No questions | Brainstorming | Unstructured interview | Structured interview | Self-report (open-ended) | Self-report (close-ended) |
| Brainstorming log | Name | ▮ | | | | | |
| | Define | ▮ | | | | | |
| | Actor | | ▮ | | | | |
| | Action | | ▮ | | | | |
| | Target | ▮ | | | | | |
| | Dose | ▮ | | | | | |
| | Temporality | | ▮ | | | | |
| | Outcome | ▮ | | | | | |
| | Justification | | | | | | |
| | Barriers | | ▮ | | | | |
| | Adaptation | | ▮ | | | | |
| Activity log | Name | ▮ | | | | | |
| | Define | ▮ | | | | | |
| | Actor | | | | | ▮ | |
| | Action | | | | | ▮ | |
| | Target | ▮ | | | | | |
| | Dose | | | | | ▮ | |
| | Temporality | | | | | ▮ | |
| | Outcome | | | | | ▮ | |
| | Justification | ▮ | | | | | |
| | Barriers | | | | | ▮ | |
| | Adaptation | ▮ | | | | | |

*(Continued)*

**Table 1.** (Continued)

| | | Not structured | | | | | Highly structured |
|---|---|---|---|---|---|---|---|
| | | No questions | Brainstorming | Unstructured interview | Structured interview | Self-report (open-ended) | Self-report (close-ended) |
| Detailed tracking log | Name | | | | | | Standardized |
| | Define | ▓ | | | | | |
| | Actor | | | | | | ▓ |
| | Action | | | | | | ▓ |
| | Target | ▓ | | | | | |
| | Dose | | | | | | ▓ |
| | Temporality | | | | | | ▓ |
| | Outcome | | | | | | ▓ |
| | Justification | ▓ | | | | | |
| | Barriers | | | | | | |
| | Adaptation | | | | | | Standardized |

Shading depicts the characteristic of each tracking tools organized by level of structure and strategy specification and additional information captured.

**Table 2.** Survey response rates.

| | Round 1 | | | Round 2 | | |
|---|---|---|---|---|---|---|
| | Activity log (%) | Brainstorming log (%) | Detailed tracking log (%) | Activity log (%) | Brainstorming log (%) | Detailed tracking log (%) |
| Mental health provider | 44.4 | 33.3 | 66.7 | 100 | 100 | 100 |
| Implementation practitioner | 83.3 | 16.7 | 66.7 | 50 | 83.3 | 83.3 |
| Treatment developer | 100 | 66.6 | 100 | 0 | 0 | 0 |

**Table 3.** Barriers and activities reported by tracking tool.

| Role | Barriers | Implementation strategies | Treatment adaptations |
|---|---|---|---|
| Brainstorming log (*N* = 24) | | | |
| Mental health provider | 7 | 1 | 8 |
| Implementation practitioner | 26 | 22 | 9 |
| Treatment developer | 2 | 1 | 3 |
| Total | 35 | 24 | 20 |
| Activity log (*N* = 17) | | | |
| Mental health provider | 3 | 2 | 1 |
| Implementation practitioner | 6 | 9 | 0 |
| Treatment developer | 0 | 1 | 0 |
| Total | 9 | 13 | 1 |
| Detailed tracking log (*N* = 22) | | | |
| Mental health provider | 1 | 7 | 3 |
| Implementation practitioner | 0 | 9 | 1 |
| Treatment developer | 1 | 3 | 0 |
| Total | 2 | 19 | 4 |

*N* represents the number of responses per log. Each response may report more than one type of activity.

log (Table 3). As shown in Figure 1, there were few significant differences between treatment adaptations reported by participants except on the reordering subscale between the activity log and the detailed tracking log. The treatment developer reported fewer treatment adaptations than the other groups (all $p$s < .05; Figure 3), but the implementation
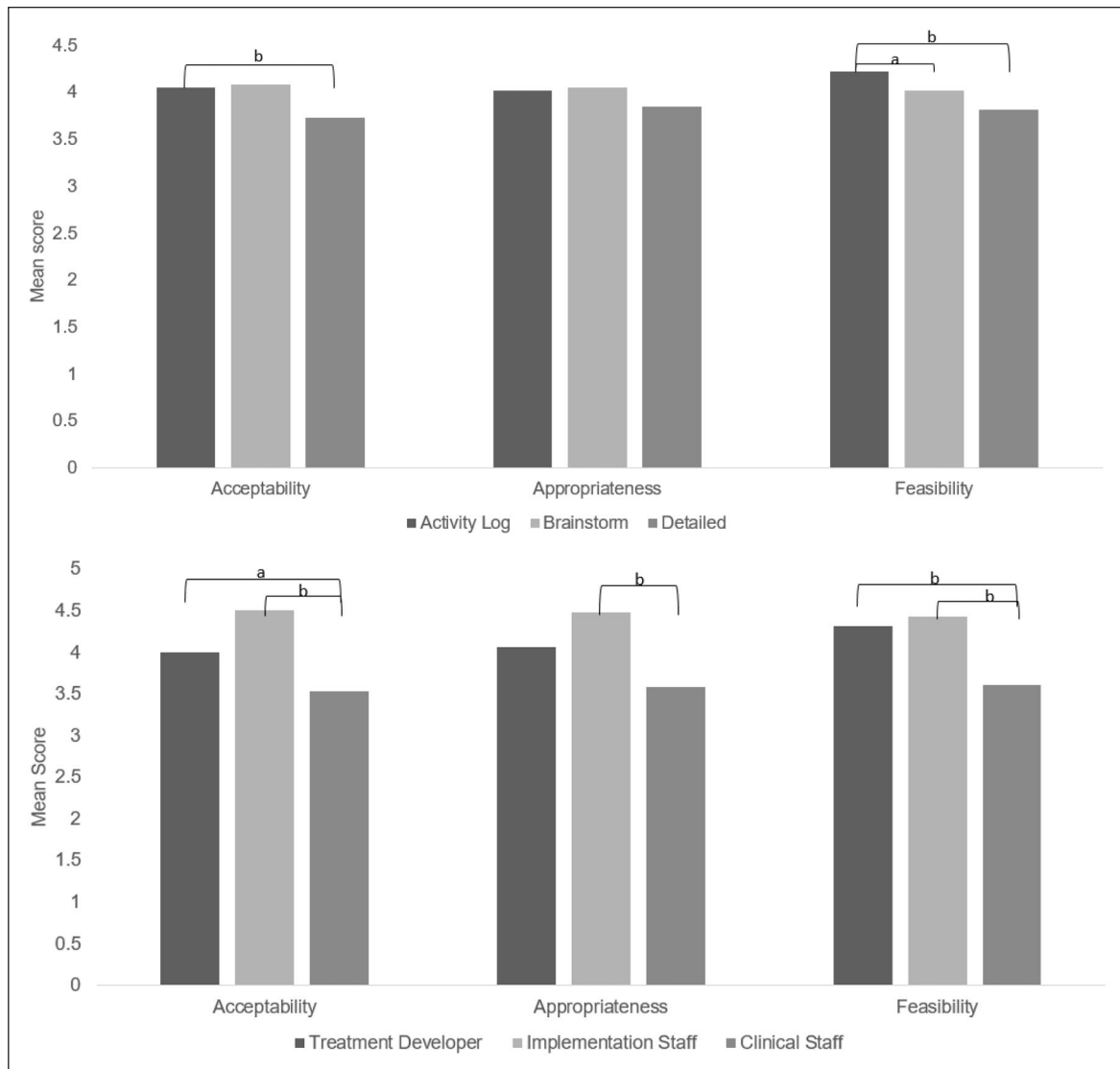
**Figure 2.** Means for acceptability of intervention measure, intervention appropriateness measure, and feasibility of intervention measure by tracking method.
a—pairwise comparison was significant at the $p < .05$ level. b—pairwise comparison was significant at the $p < .01$ level.

practitioners and mental health providers did not significantly differ on reporting of adaptations.

## Qualitative findings

*Tracking process.* The average completion time across all three tools was less than 15 min. The activity log had the longest completion time ($M = 14$ min), while the detailed tracking log was the shortest ($M = 8.5$ min). The desire for brevity was expressed across all participants. One mental health provider noted that the ideal length of time to complete tracking is "ten minutes . . . in ten minutes you can do a lot of work." Another provider indicated "I have a lot of families . . . to follow through with. The fact that it was not a long survey . . . was great."

All participants agreed internet was the preferred mode of accessing the tools. One implementation practitioner recalled, "I thought that was super simple. I had a reminder in my calendar each week to do it." Although most participants indicated that weekly tracking was reasonable, there was some concern regarding frequency. The treatment developer reflected, "I don't know if it makes sense to wait until an entire Blues Program group is done. I don't think they would remember the changes that were made . . . [but] somehow weekly seems too much." Indeed, several providers noted time and competing priorities as the largest barriers to tracking completion. One provider reported,

We have children's plan progress notes, administrative things to do, . . . ongoing trainings, . . . staff meetings, . . . crises . . .
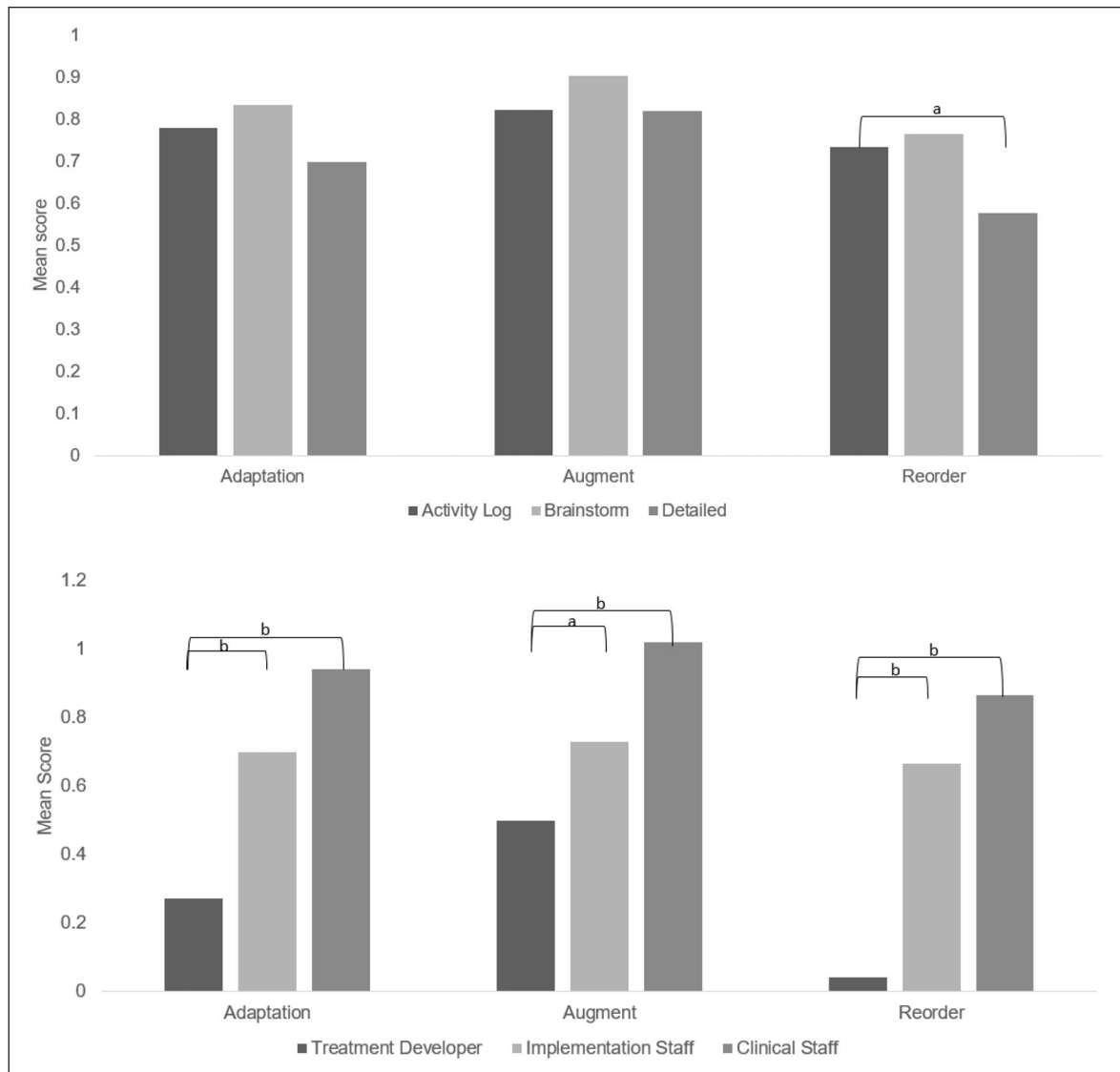
**Figure 3.** Means for the adaptation of EBP scale.
a—pairwise comparison was significant at the $p < .05$ level. b—pairwise comparison was significant at the $p < .01$ level.

to attend to that we can't account for . . . that could prevent it [tracking] from being completed on time . . . I would prefer it to be biweekly.

There was variation in how the tools were interpreted by participants and the types of activities reported on each. One provider reflected, ". . . my coworkers . . . had different answers or different ways of viewing the questions . . . I just thought that was interesting how we all had very different perspectives, and different feedback to give . . ." An implementation practitioner indicated "with the Brainstorming Log . . . at times I connected it back a lot to the actual facilitation of the group . . . Because I wasn't facilitating at that time, I probably . . . gave a lot more information in the Detailed Tracking Logs." In general, providers reported information on Blues Program sessions, whereas the

treatment developer and implementation practitioners reported activities, such as training, communication, preparation to implement the Blues Program at a new school, and other activities peripheral to EBP delivery.

*Brainstorming log.* Participants indicated the brainstorming log was easy to use and provided an opportunity to reflect on recent activities. The treatment developer preferred this tool, stating ". . . it's helpful when you're defining the content of the Blues and then the context and barriers is really clear, and the strategies, I like this one the best." Several participants reacted positively to the space to describe barriers and the open-ended questions. However, some mental health providers indicated a preference for multiple choice questions; one suggested using multiple choice with an option to provide additional narrative. Several providers

felt there was a lack of clarity around the types of activities that should be reported, with one provider indicating some activities (e.g., providing "goodie bags" to students) were not captured by the questions.

*Activity log.* Participants stated an overall strength of the activity log was its ease of use and indicated the questions provided enough structure and detail without being overly restrictive. An implementation practitioner indicated, "you're orienting it that way with purpose and outcome help to understand the intentionality behind what you're doing. So, I appreciated that about this one." Several participants noted the activity log was a helpful planning tool. A provider reflected,

> how much time you spent on it, that will allow me to plan ahead in regard to executing the program. Why did you choose this activity . . . what was the outcome . . . how would you know if it actually was effective . . . or did it do what it was supposed to do? That will be great information for me.

As with the brainstorming log, several participants perceived a lack of clarity in the activity log. The "date" prompt created confusion for several providers, as they were uncertain if they should report calendar dates or Blues Program session number. One provider noted, ". . . initially it seemed the easiest. But it actually is kind of confusing. I wasn't quite sure how to chunk up the actual problem or process I was going to write about."

*Detailed tracking log.* Several participants indicated that the ability to select a pre-determined option was the strength of the detailed tracking log. One provider stated, "I liked that because then you were able to . . . narrow it down." The treatment developer indicated that the definitions provided for implementation strategies and treatment adaptations were helpful. The sentiment reflected by a provider, "I don't know if it was just not relevant or if it was just the wording that didn't make it seem relevant," was expressed by several participants when describing the strategy labels. Another provider noted the participant burden,

> it's a lot of work . . . they're giving you options to help whoever is taking their survey compartmentalize what they did or how they adjusted it. Giving them . . . wording for it is helpful, or categories for it. But it is a lot of clicking around, and that might get confusing.

The treatment developer indicated difficulty in applying implementation science language, noting,

> This is one where I needed to be educated . . . I was willing to try to read it and think about what it might have meant. But I did remember thinking . . . if the facilitators read this, they don't have the time . . . to try to understand.

*Recommendations for process improvement.* All participants indicated that additional background and training in the tracking process would have been useful. An implementation practitioner reflected,

> When I would send them [surveys] out, I did give a lot of rationale to the clinicians . . . Had I been able to look at them ahead of time, and really talk to you guys a little bit more about how they're going to be done, and questions that could come up by our frontline staff . . . to be able to better support them understanding the rationale, and what the differences are between implementation strategy and treatment adaptation. I know that they're not going to understand everything, nor am I . . . had that opportunity it might have been a little bit easier for them [mental health providers].

Most participants indicated desire for an introduction to tracking during the Blues Program training. Several participants noted that there was a learning curve associated with tracking and with distinguishing between implementation strategies and treatment adaptations. One implementation practitioner recalled,

> in the beginning, I was still trying to figure that out. But I think over the past six months, I definitely see the distinction and I think if I were to have another opportunity doing this, I'd probably look at it a little differently.

There was variation not only in how the tools were rated but also how participants interpreted the tracking tools. An implementation practitioner noted of the brainstorming log, "this kind of more free text boxes might have made more sense as. . . a post-implementation strategy survey." Several participants indicated that they would be more likely to engage in tracking not only if they had a better understanding of the expectations but also how tracking related to their work. One provider noted,

> I know it would be helpful for those coordinating the program, but if I knew it was for my benefit as well, it would motivate me more, definitely. To know I could reflect on it going into the next group I'm facilitating and seeing what adaptations I did or what may have worked or not.

When asked how tracking could be more useful, one provider suggested, "it would probably be our supervisor to go through the feedback regarding the surveys."

## Discussion

Given the importance of planned implementation strategy tailoring and treatment adaptation to promote the sustainability of EBPs (Chambers et al., 2013), we piloted three tools in a mixed-methods study designed to prospectively track implementation strategies and treatment adaptations. This work builds on previous efforts to develop tracking tools and coding systems to capture information about

implementation strategy planning and tailoring efforts (Boyd et al., 2017; Bunger et al., 2017; Dogherty et al., 2012; Ritchie et al., 2020), determine which tool is most feasible, acceptable, and appropriate to users, and which yields actionable data that aligns with reporting recommendations (Proctor et al., 2013). Our study offers preliminary reliability, validity, and pragmatic utility evidence across three tools that attempt to disentangle implementation strategy tailoring (Rudd et al., 2020; Wilson et al., 2017) and treatment adaptation (Baumann et al., 2017; Wiltsey Stirman et al., 2013). While others have noted difficulty in distinguishing between treatments and implementation strategies in tracking (Rabin et al., 2018), this is a necessary and worthwhile endeavor that can benefit research and practice (Eldh et al., 2017; National Academies of Sciences Engineering and Medicine; Division of Behavioral and Social Sciences and Education; Board on Children, Youth, and Families; Committee on Fostering Healthy Mental, Emotional, and Behavioral Development Among Children and Youth, 2019).

Participants perceived the tools as acceptable, appropriate, and feasible, although ratings were poorer among mental health providers. Notably, this was the only group not involved in tool development, which may indicate provider needs were not met and those of other participants. As Glasgow and Riley (2013) note, priorities often vary across stakeholders and they should be involved early and throughout the development process to ensure relevance and actionability. All participants offered useful feedback on formatting, content, administration process, and the ultimate purpose and function of the tools. As we refine these tools for further testing, we will seek additional opportunities for input from a more inclusive array of stakeholders across a range of settings.

Participants rated the activity log as the most feasible, and more acceptable than the detailed tracking log. However, the brainstorming log revealed the most discrete activities and barriers. Furthermore, the activity log revealed only one treatment adaptation compared to 20 in the brainstorming log (Table 3), even though AES scores from the two tools were similar (Figure 3). This is surprising, as several providers indicated they utilized the tracking tools to report on their Blues Program sessions. Even though the activity log was rated higher, it may not be the most effective tool for eliciting actionable data, especially treatment adaptations. This could be due to the lack of instruction to differentiate between implementation strategies and treatment adaptations as the other two tools had. Given the importance of tracking, further testing is needed to detect reporting differences when these instructions are present or absent. The lack of alignment between tool preference and the actionability of data highlights the trade-offs associated with tool structure previously mentioned. Considering the potential for adaptations made by front-line providers to improve EBP fit with their context

(Chambers & Norton, 2016; Lau, 2006), having a tool that elicits detailed, reliable information about these adaptations is crucial.

It is important to note that all participants found this tracking to be quite different from anything they had previously done. The implementation practitioners gave the tracking tools the highest rating of the three groups and had the most positive perceptions. Though all groups reported tracking had utility, this was least salient among mental health providers. Furthermore, providers most frequently perceived a lack of clarity when tracking their activities. This suggests the need to involve all stakeholder groups earlier in the process, and that tracking may be less meaningful if approached only as a data collection activity. Indeed, several providers noted they would be more likely to track their activities if they could connect tracking to service delivery and if they could review their data and receive feedback from their supervisor. Stakeholders' needs and the utility of tracking are important considerations in tool design and administration (Glasgow & Riley, 2013).

The variability in how the tools were interpreted and the types of activities reported on each may indicate the need or desire for different versions of tracking tools at various stages of implementation. Most providers treated tracking as an accompaniment to their session notes, and the content of their tools focused on the delivery of the group sessions. This may suggest an important opportunity, especially given that several providers noted it was helpful to think prospectively about barriers and strategies to overcome them. Given the brainstorming log elicited the most barriers, it seems this tool would be best suited for this purpose. However, an implementation practitioner noted that this tool seemed like a post-implementation strategy survey. Further testing should reveal if different tracking tools elicit different kinds of data at various implementation phases and if the nature of the role (e.g., delivering services vs. overseeing implementation) necessitates a unique kind of reporting. Understanding these nuances will better enable the field to design tools that allow for capturing complete, detailed information about implementation strategies (Powell et al., 2019; Proctor et al., 2013).

All participants noted that additional background and training about tracking would be helpful, and agreed that this information would be best delivered during training in the EBP; however, one-time training would not be sufficient. Providers most frequently indicated a desire for follow-up, whether from implementation practitioners or clinical supervisors. Implementation practitioners, though engaged throughout the design and data collection process, also indicated a desire for additional opportunities to discuss tracking and receive more information. This illustrates that tracking is not a "set it and forget it" activity, rather is dynamic and requires iterative feedback.

While this pilot made strides toward developing tools to prospectively track treatment adaptations and implementation strategies in a low-burden manner, there were several limitations. First, due to the scale of the Blues Program implementation, the sample size was small, with 11 participants across three roles. The analyses of acceptability, appropriateness, feasibility, and adaptation by tool and role should be considered preliminary and replicated in larger samples. In addition, the survey response rate varied week to week and across participant groups, ranging from 0% to 100%. Two main factors affected the response rates: (1) a program champion was appointed prior to the second round, which improved provider engagement, and (2) the treatment developer did not conduct training during the second data collection round, thus did not have new activities to report and did not complete tracking. The nature of the Blues Program may limit generalizability of our findings to other treatment contexts. The group delivery of the EBP in school contexts poses implementation challenges that may be substantially different than those for individual treatments in clinical contexts or large-scale public health interventions in community settings. Finally, most participants were unfamiliar with implementation science terminology and identified needs for additional training, instructions, and examples to improve clarity. It also suggests a tension between the importance of reporting implementation strategies using consistent terminology and the need to conduct tracking in a user-friendly way using terminology that might be clearer or more easily understood. Future work will serve to refine the tracking methods and test them with a larger group of stakeholders in various contexts.

## Conclusion

This study sought to fill methodological gaps that prevent stakeholders and researchers from discerning which activities are most important for promoting EBP sustainment. Tracking is essential for implementation research to understand how to build effective implementation strategies and replicate effects. Tracking is also beneficial for implementation practice and clinical practice as it allows practitioners to better tailor efforts to their local context and to ensure EBPs are delivered with fidelity. Three tools show promise for these purposes, but additional work is needed to assess their psychometric and pragmatic properties and ensure they serve their intended purpose. Ultimately, these tools could inform prospective tailoring of implementation strategies and treatment adaptations, which would promote EBP sustainment.

## Declaration of conflicting interests

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

## ORCID iDs

Callie Walsh-Bailey [iD] https://orcid.org/0000-0002-1417-5130

Kayne D Mettert [iD] https://orcid.org/0000-0003-1750-7863

Byron J Powell [iD] https://orcid.org/0000-0001-5245-1186

Shannon Wiltsey Stirman [iD] https://orcid.org/0000-0001-9917-5078

Aaron R Lyon [iD] https://orcid.org/0000-0003-3657-5060

Cara C Lewis [iD] https://orcid.org/0000-0001-8920-8075

## Supplemental material

Supplemental material for this article is available online.

## References

Albright, L., & Cobb, B. R. (1988, April 5–9). *Formative evaluation of a training curriculum for vocational education and special services personnel* [Paper presentation]. Annual Meeting of the American Educational Research Association, New Orleans, LA, United States.

Baker, R., Camosso-Stefinovic, J., Gillies, C., Shaw, E. J., Cheater, F., Flottorp, S., Robertson, N., Wensing, M., Fiander, M., Eccles, M. P., Godycki-Cwirko, M., van Lieshout, J., & Jäger, C. (2015). Tailored interventions to address determinants of practice. *Cochrane Database of Systematic Reviews*, *4*, CD005470. https://doi.org/10.1002/14651858.CD005470.pub3

Baumann, A. A., Cabassa, L. J., & Wiltsey Stirman, S. (2017). Adaptation in dissemination and implementation science. In R. C. Brownson, G. A. Colditz, & E. K. Proctor (Eds.), *Dissemination and implementation research in health: Translating science to practice* (pp. 285–300). Oxford University Press. https://doi.org/10.1093/oso/9780190683214.001.0001

Boyd, M. R., Powell, B. J., Endicott, D., & Lewis, C. C. (2017). A method for tracking implementation strategies: An exemplar implementing measurement-based care in community behavioral health clinics. *Behavior Therapy*, *49*(4), 525–537. https://doi.org/10.1016/j.beth.2017.11.012

Bunger, A. C., Powell, B. J., Robertson, H. A., MacDowell, H., Birken, S. A., & Shea, C. (2017). Tracking implementation strategies: A description of a practical approach and early findings. *Health Research Policy and Systems*, *15*(1), 15. https://doi.org/10.1186/s12961-017-0175-y

Chambers, D. A., Glasgow, R. E., & Stange, K. C. (2013). The dynamic sustainability framework: Addressing the paradox of sustainment amid ongoing change. *Implementation Science*, *8*, 117. https://doi.org/10.1186/1748-5908-8-117

Chambers, D. A., & Norton, W. E. (2016). The adaptome: Advancing the science of intervention adaptation. *American Journal of Preventive Medicine*, *51*(4, Suppl. 2), S124–S131. https://doi.org/10.1016/j.amepre.2016.05.011

Cook, C. R., Lyon, A. R., Locke, J., Waltz, T., & Powell, B. J. (2019). Adapting a compilation of implementation strategies to advance school-based implementation research and practice. *Prevention Science*, *20*, 914–935. https://doi.org/10.1007/s11121-019-01017-1

Demetriou, C., Ozer, B. U., & Essau, C. A. (2015). Self-report questionnaires. In R. L. Cautin & S. O. Lilienfeld (Eds.), *The encyclopedia of clinical psychology* (pp. 1–6). John Wiley.

Dogherty, E. J., Harrison, M. B., Baker, C., & Graham, I. D. (2012). Following a natural experiment of guideline adaptation and early implementation: A mixed-methods study of facilitation. *Implementation Science*, *7*(1), 9. https://doi.org/10.1186/1748-5908-7-9

Edwards, P. (2010). Questionnaires in clinical trials: Guidelines for optimal design and administration. *Trials*, *11*, 2. https://doi.org/10.1186/1745-6215-11-2

Eldh, A. C., Almost, J., DeCorby-Watson, K., Gifford, W., Harvey, G., Hasson, H., Kenny, D., Moodie, S., Wallin, L., & Yost, J. (2017). Clinical interventions, implementation interventions, and the potential greyness in between—A discussion paper. *BMC Health Services Research*, *17*(1), Article 16. https://doi.org/10.1186/s12913-016-1958-5

Fixsen, D., Blase, K., Metz, A., & Van Dyke, M. (2013). Statewide implementation of evidence-based programs. *Exceptional Children*, *79*(3), 213–230. https://doi.org/10.1177/001440291307900206

Glasgow, R. E., & Riley, W. T. (2013). Pragmatic measures: What they are and why we need them. *American Journal of Preventive Medicine*, *45*(2), 237–243. https://doi.org/10.1016/j.amepre.2013.03.010

Hill, C. E., Knox, S., Thompson, B. J., Williams, E. N., Hess, S. A., & Ladany, N. (2005). Consensual qualitative research: An update. *Journal of Counseling Psychology*, *52*(2), 196–205. https://doi.org/10.1037/0022-0167.52.2.196

Kilbourne, A. M., Almirall, D., Eisenberg, D., Waxmonsky, J., Goodrich, D. E., Fortney, J. C., Kirchner, J. E., Solberg, L. I., Main, D., Bauer, M. S., Kyle, J., Murphy, S. A., Nord, K. M., & Thomas, M. R. (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*(1), 132. https://doi.org/10.1186/s13012-014-0132-x

Kirchner, J. E., Waltz, T. J., Powell, B. J., Smith, J. L., & Proctor, E. K. (2017). Implementation strategies. In R. C. Brownson, G. A. Colditz, & E. K. Proctor (Eds.), *Dissemination and implementation research in health: Translating science to practice* (Vol. 2, pp. 245–266). Oxford University Press.

Lau, A. S. (2006). Making the case for selective and directed cultural adaptations of evidence-based treatments: Examples from parent training. *Clinical Psychology & Practice*, *13*, 295–310. https://doi.org/10.1111/j.1468-2850.2006.00042.x

Lau, A. S., Barnett, M., Stadnick, N., Saifan, D., Regan, J., Wiltsey Stirman, S., Roesch, S., & Brookman-Frazee, L. (2017). Therapist report of adaptations to delivery of evidence-based practices within a system-driven reform of publicly funded children's mental health services. *Journal of Consulting and Clinical Psychology*, *85*(7), 664–675. https://doi.org/10.1037/ccp0000215

Lewis, C. C., Scott, K., Marti, C. N., Marriott, B. R., Kroenke, K., Putz, J. W., Mendel, P., & Rutkowski, D. (2015). Implementing Measurement-Based Care (iMBC) for depression in community mental health: A dynamic cluster randomized trial study protocol. *Implementation Science*, *10*, 127. https://doi.org/10.1186/s13012-015-0313-2

National Academies of Sciences Engineering and Medicine; Division of Behavioral and Social Sciences and Education; Board on Children, Youth, and Families; Committee on Fostering Healthy Mental, Emotional, and Behavioral Development Among Children and Youth. (2019). Effective implementation: Core components, adaptation, and strategies. In *Fostering healthy mental, emotional, and behavioral development in children and youth: A national agenda*. National Academies Press. https://www.ncbi.nlm.nih.gov/books/NBK551850/

Powell, B. J., Beidas, R. S., Lewis, C. C., Aarons, G. A., McMillen, J. C., Proctor, E. K., & Mandell, D. S. (2017). Methods to improve the selection and tailoring of implementation strategies. *The Journal of Behavioral Health Services & Research*, *44*(2), 177–194. https://doi.org/10.1007/s11414-015-9475-6

Powell, B. J., Fernandez, M. E., Williams, N. J., Aarons, G. A., Beidas, R. S., Lewis, C. C., McHugh, S. M., & 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., 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*, 21. 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*, 139. https://doi.org/10.1186/1748-5908-8-139

Rabin, B. A., McCreight, M., Battaglia, C., Ayele, R., Burke, R. E., Hess, P. L., Frank, J. W., & Glasgow, R. E. (2018). Systematic, multimethod assessment of adaptations across four diverse health systems interventions. *Frontiers in Public Health*, *6*, 102. https://doi.org/10.3389/fpubh.2018.00102

Reja, U., Manfreda, K. L., Hlebec, V., & Vehovar, V. (2003). Open-ended vs. close-ended questions in web questionnaires. *Advances in Methodology and Statistics*, *19*(1), 106–117.

Ringmayr, T. G. (2013). *ATLAS.ti* (Version 7.1.0). ATLAS.ti Scientific Software Development GmbH.

Ritchie, M. J., Kirchner, J. E., Townsend, J. C., Pitcock, J. A., Dollar, K. M., & Liu, C.-F. (2020). Time and organizational cost for facilitating implementation of primary care mental health integration. *Journal of General Internal Medicine*, *35*(4), 1001–1010. https://doi.org/10.1007/s11606-019-05537-y

Rudd, B. N., Davis, M., & Beidas, R. S. (2020). Integrating implementation science in clinical research to maximize

public health impact: A call for the reporting and alignment of implementation strategy use with implementation outcomes in clinical research. *Implementation Science*, *15*(1), 103. https://doi.org/10.1186/s13012-020-01060-5

Stice, E., Rohde, P., Gau, J. M., & Wade, E. (2010). Efficacy trial of a brief cognitive-behavioral depression prevention program for high-risk adolescents: Effects at 1- and 2-year follow-up. *Journal of Consulting and Clinical Psychology*, *78*(6), 856–867. https://doi.org/10.1037/a0020544

Stice, E., Rohde, P., Seeley, J. R., & Gau, J. M. (2008). Brief cognitive-behavioral depression prevention program for high-risk adolescents outperforms two alternative interventions: A randomized efficacy trial. *Journal of Consulting and Clinical Psychology*, *76*(4), 595–606. https://doi.org/10.1037/a0012645

Tuckett, A. G., & Stewart, D. E. (2004). Collecting qualitative data: Part II journal as a method: Experience, rationale and limitations. *Contemporary Nurse*, *16*(3), 240–251. https://doi.org/10.5172/conu.16.3.240

Waltz, T. J., Powell, B. J., Matthieu, M. M., Damschroder, L. J., Chinman, M. J., Smith, J. L., Proctor, E. K., & Kirchner, J. E. (2015). Use of concept mapping to characterize relationships among implementation strategies and assess their feasibility and importance: Results from the Expert Recommendations for Implementing Change (ERIC) study. *Implementation Science*, *10*, 109. https://doi.org/10.1186/s13012-015-0295-0

Weiner, B. J., Lewis, C. C., Stanick, C., Powell, B. J., Dorsey, C. N., Clary, A. S., Boynton, M. H., & Halko, H. (2017). Psychometric assessment of three newly developed implementation outcome measures. *Implementation Science*, *12*(1), 108. https://doi.org/10.1186/s13012-017-0635-3

Wensing, M. (2017). The Tailored Implementation in Chronic Diseases (TICD) project: Introduction and main findings. *Implementation Science*, *12*(1), 5. https://doi.org/10.1186/s13012-016-0536-x

Wilson, P. M., Sales, A., Wensing, M., Aarons, G. A., Flottorp, S., Glidewell, L., Hutchinson, A., Presseau, J., Rogers, A., Sevdalis, N., Squires, J., & Straus, S. (2017). Enhancing the reporting of implementation research. *Implementation Science*, *12*(1), 13. https://doi.org/10.1186/s13012-017-0546-3

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*(1), 58. https://doi.org/10.1186/s13012-019-0898-y

Wiltsey Stirman, S., Finley, E. P., Shields, N., Cook, J., Haine-Schlagel, R., Burgess, J. F., Jr., Dimeff, L., Koerner, K., Suvak, M., Gutner, C. A., Gagnon, D., Masina, T., Beristianos, M., Mallard, K., Ramirez, V., & Monson, C. (2017). Improving and sustaining delivery of CPT for PTSD in mental health systems: A cluster randomized trial. *Implementation Science*, *12*(1), 32. https://doi.org/10.1186/s13012-017-0544-5

Wiltsey Stirman, S., Miller, C. J., Toder, K., & Calloway, A. (2013). Development of a framework and coding system for modifications and adaptations of evidence-based interventions. *Implementation Science*, *8*(1), 65. https://doi.org/10.1186/1748-5908-8-65