A randomized controlled trial testing supervision strategies in community mental health

Shannon Dorsey[1] , Rashed AlRasheed[1], Suzanne EU Kerns[2] ,
Rosemary D Meza[3], Noah Triplett[1] , Esther Deblinger[4],
Nathaniel Jungbluth[5], Lucy Berliner[6,7,8], Lavangi Naithani[9]
and Michael D Pullmann[8]

Abstract

Background

Clinicians need supports beyond training to deliver evidence-based treatments with fidelity. Workplace-based clinical supervision often is a commonly provided support in community mental health, yet too few studies have empirically examined supervision and its impact on clinician fidelity and treatment delivery.

Method

Building on a Washington State-funded evidence-based treatment initiative (CBT+), we conducted a randomized controlled trial (RCT), testing two supervision conditions delivered by workplace-based supervisors (supervisors employed by community mental health organizations). The RCT followed a supervision-as-usual (SAU) phase for comparison. The treatment of focus was trauma-focused cognitive behavioral therapy (TF-CBT). Clinicians ($N =$ 238) from 25 organizations participated in the study across the SAU baseline and RCT phases. In the RCT phase, clinicians were randomized to either symptom and fidelity monitoring (SFM) or SFM and behavioral rehearsal (SFM + BR). For BR, clinicians engaged in a short role play of an upcoming treatment element. Supervisors delivered both conditions, with regular study monitoring for drift. Clinicians audiorecorded therapy sessions with enrolled clients, and masked coders coded a subset of recordings for adherence to TF-CBT. One hundred and thirty-three clinicians had recorded TF-CBT session data for 258 youth. We examined six adherence outcomes, including potential moderators.

Results

Results of generalized estimating equations indicated that there were no real differences on adherence outcomes for experimental conditions (SFM, SFM + BR) compared to SAU. Adherence scores in the baseline SAU phase and the RCT conditions were high. Only one interaction was significant.

Conclusions

Contrary to our hypotheses, we did not see improvements in adherence with the RCT conditions. However, nonsignificant findings seem best explained by clinicians' acceptable/high adherence in SAU. This study was

[1]Department of Psychology, University of Washington, Seattle, WA, USA
[2]The Kempe Center, University of Colorado—Anschutz, Aurora, CO, USA
[3]Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA
[4]Child Abuse Research Education and Service (CARES) Institute, Rowan Medicine, Stratford, NJ, USA
[5]Seattle Children's Hospital, Seattle, WA, USA
[6]Harborview Center for Sexual Assault and Traumatic Stress, University of Washington, Seattle, WA, USA
[7]School of Social Work, University of Washington, Seattle, WA, USA
[8]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA
[9]School/Clinical Psychology, The University of Texas at Austin, Austin, TX, USA

Corresponding author:
Shannon Dorsey, Department of Psychology, University of Washington, Kincaid Hall, Seattle, WA 98195, USA.
Email: dorsey2@uw.edu

conducted within the context of a long-standing, state-funded EBT initiative, in which clinicians and their supervisors receive training and support, and in which participating community mental health organizations have adopted and supported TF-CBT.

Plain Language Summary
Clinicians often receive training in evidence-based treatments (EBTs), but training alone is not enough for clinicians to deliver EBTs well. They need other supports. One of the necessary supports is clinical supervision. In public community mental health organizations, which are often resource-constrained, routine clinical supervision is a commonly available support. However, its potential for supporting high-quality EBT delivery is not well understood. This study looked at supervision and clinicians' delivery of trauma-focused cognitive behavioral therapy (TF-CBT), an EBT for youth who have mental health problems subsequent to trauma. We found that clinicians who received workplace-based supervision on TF-CBT from TF-CBT-trained supervisors, within the context of a state-funded EBT initiative, delivered TF-CBT well. This initiative supported delivery by providing training and supervisor-specific supports. In a second phase of the study, we rigorously tested supervisors' use of two specific technique packages. These packages included having supervisors review: (a) graphed client mental health symptoms across treatment and clinicians' report of TF-CBT elements delivered each session or (b) these *plus* having clinicians do a brief demonstration of how they would deliver an upcoming TF-CBT element to their client, so supervisors could "see" clinicians' skill and provide feedback. Contrary to expectations, we did not see important differences in clinicians' TF-CBT delivery. This may be because the clinicians enrolled in our study were already delivering TF-CBT well and, therefore, there was less opportunity for changes. Future studies should explore if these techniques are more helpful for specific situations or clinicians.

Introduction

Mental health-related concerns are prevalent among youth in the U.S. (Bitsko et al., 2022). Because most youth with mental health needs do not receive treatment (Bitsko et al., 2022; Whitney & Peterson, 2019) and because evidence-based treatments (EBT) are not always available in community-settings (Bruns et al., 2016), many states and local jurisdictions are interested in increasing the availability of EBTs for youth and families (e.g., Ashwood et al., 2018; Beidas et al., 2019). However, literature documents a range of challenges. One challenge involves delivering EBTs with fidelity, a multidimensional construct defined broadly as the extent to which treatment is delivered as intended (Tabak et al., 2012). To deliver EBTs with fidelity and be responsive to client needs, providers need support in appropriately tailoring treatment to their clients (Kendall & Frank, 2018) and the local context (Beidas et al., 2019). Identification of support strategies that are readily available within community mental health centers (CMHCs) is essential for high-quality mental health availability (Glasgow, 2013).

Workplace-based clinical supervision is a commonly provided support in CMHCs and has been highlighted as a potentially cost-neutral and feasible implementation strategy for EBT implementation (Schoenwald et al., 2008; Tugendrajch et al., 2023) that can improve EBT fidelity (Beidas & Kendall, 2010; Henggeler et al., 2002)

and client outcomes (Bambling et al., 2006; Schoenwald et al., 2009). In some studies, supervision may matter as much, or more, than the modality of training for adherence and competency in EBTs (Beidas et al., 2012; Rakovshik et al., 2016; Sholomskas et al., 2005).

Theoretical work on supervision (Milne & Reiser, 2012) and clinician training (McLeod et al., 2018b) highlight several "gold standard" EBT supervision techniques (Milne, 2007; Roth et al., 2010). These include opportunities for practice, skill-building, and feedback which can activate learning processes that theoretically facilitate learning and reflection (Beidas et al., 2014; Kolb, 1984). Specifically, these supervision techniques include behavioral rehearsal, live or recorded session review, fidelity monitoring, and client symptom monitoring.

While workplace-based supervision research in CMHCs is limited, the available studies provide some empirical support for these techniques. An observational study of multisystemic therapy (MST) found that workplace supervisors' use of active practice of skills, case conceptualization, and monitoring fidelity to MST, predicted greater clinician MST adherence and better client outcomes (Schoenwald et al., 2009). A more recent experimental study found that a supervision approach emphasizing corrective feedback based on actual practice review, and skills coaching using behavioral rehearsal, led to greater competency in an EBT than supervision as usual (Martino et al., 2016), with

behavioral rehearsal defined as a simulated interaction, or role play, involving a trainee and another individual (Cross et al., 2007). Similar results were obtained in a study focused on graduate student trainees (Bearman et al., 2017). These findings, combined with related research focused on EBT expert consultation (Choy-Brown & Stanhope, 2018; Miller et al., 2004), provide empirical support for supervision techniques involving practice, observation, feedback, and progress and fidelity monitoring.

While the literature provides evidence for techniques that may impact treatment delivery, studies empirically testing specific supervision techniques in CMHCs are needed. Further, we lack an understanding of which techniques are most beneficial. This understanding is essential, given that although many organizations provide regular supervision for clinical providers, the typical 1–2 hours weekly is limited considering high caseloads (Choy-Brown & Stanhope, 2018; Dorsey, Pullmann et al., 2017). Additionally, workplace-based supervisors use supervision to accomplish many competing functions (Dorsey, Pullmann et al., 2017), including emotional support to clinicians (Fukui et al., 2019) and covering administrative needs. As such, guidance on how to maximize potential impact of supervision is needed.

Another level of precision largely absent from supervision research is what works for whom. The effectiveness of supervision techniques in improving clinician adherence may differ (i.e., be moderated) by clinician or supervisor characteristics. Social cognitive theory posits that self-efficacy is important for behavior (Bandura, 1998) and has been linked to successful implementation of physical activity interventions (Huijg et al., 2013). Both self-efficacy and knowledge play important theoretical roles in a mechanistic model of clinician training and consultation (McLeod et al., 2018a). For supervisors, their self-efficacy in supervising an EBT, or EBT knowledge, may affect supervision technique delivery, impacting clinicians' adherence. For clinicians, self-efficacy or knowledge in the EBT may impact benefit of specific supervision techniques. Supervisors may adjust their approach based on clinician characteristics (Milne, 2009; Milne & James, 2000), as found in a study of workplace-based supervisors where a more directive approach was used with clincians less experienced in the EBT (Meza et al., 2023). In a study using direct observation of EBT-focused supervision, supervisors with less EBT knowledge were more likely to cover non-EBT content (Pullmann et al., 2018).

This article presents the findings from a within and between subjects hybrid type 3 randomized controlled trial (RCT) of supervision strategies, conducted in publicly funded CMHCs. The RCT compares clinician adherence, an important aspect of fidelity (Dorsey et al., 2013) for two randomized supervision conditions that followed a baseline supervision-as-usual (SAU) phase. The randomized conditions are (a) symptom and fidelity monitoring (SFM) and (b) SFM plus behavioral rehearsal (SFM + BR). We hypothesized that adherence would be highest for SFM + BR, given that the BR involved an active, experiential activity for clinicians and provided supervisors the opportunity to observe an "analog" of clinicians' treatment delivery (Beidas et al., 2014). We hypothesized that both SFM and SFM + BR would outperform SAU, as both experimental conditions include opportunities for reflection on treatment delivery and client response. Clinicians delivered trauma-focused cognitive behavioral therapy (TF-CBT) (Cohen et al., 2017), an EBT for youth mental health symptoms subsequent to trauma exposure. Given that some implementation strategies, like some clinical interventions, may work best for subsets of individuals, we also examined potential moderators with empirical support in the literature, including clinicians' objective knowledge and self-efficacy in TF-CBT and supervisors' TF-objective knowledge in TF-CBT and self-efficacy in supervising TF-CBT.

Methods

Data came from the NIMH-funded *Supervision to Enhance Practice Study (STEPS)* focused on workplace-based clinical supervision of TF-CBT (Dorsey et al., 2013), with primary aims of (a) describing "baseline" SAU strategies (Phase I; September 2012–August 2013) and (b) evaluating the effects of two different supervision conditions relative to each other and to baseline SAU on clinician adherence and client outcomes (Phase II; September 2013–December 2016). The Phase II supervision conditions incorporate gold standard elements from efficacy and effectiveness trials.

Procedures

Ethical Review. Washington State IRB, and subsequently the University of Washington IRB, approved all study procedures.

Study Sample Recruitment. The present study builds on a longstanding academic-community partnership, the Washington State EBT initiative (called CBT+), focused on public mental health (Dorsey et al., 2016). In a 2015 analysis, 83% of all public mental health organizations in Washington had participated. The CBT+ initiative provided training in TF-CBT, as well as depression, anxiety, and behavior problems, for CMHCs. Organizations participating in CBT+ sent at least one supervisor for 3 days of in-person training and 6 months of twice-monthly consultation, focused on applying the training to cases. Supervisors had the option to join technical assistance calls every month and a yearly supervisor training. To address trained provider attrition, organizations could send clinicians each year.

Organizations that participated in CBT+ and still had at least one TF-CBT-trained supervisor were approached. Our planned sample size targets were determined via an a priori power analysis. Detailed study descriptions were provided to senior leaders and potentially eligible supervisors, followed by informed consent procedures with

interested supervisors. Across Phases I and II, of the 33 organizations approached, 25 (76%) were interested in participating. Enrolled supervisors identified potentially eligible clinicians. The research team invited them to participate, screened for eligibility, and obtained informed consent. Enrolled clinicians introduced the study to guardians of youth who met study criteria. If families expressed interest, study staff conducted final eligibility screening, and informed consent/assent. Clinicians audiorecorded all TF-CBT sessions with enrolled study youth for 6 months or until treatment termination, whichever happened first. Clinicians labeled recordings with the primary treatment component for that session and recordings were automatically date-stamped. Any portion of supervision meetings in which enrolled TF-CBT cases were discussed were also recorded. At the end of the Phase II RCT, organizations received $3,000 for participation. Compensation for completing surveys was $30 for supervisors and clinicians, $20 for guardians, and $10 for youth.

In Phase II, clinicians continuing from Phase I and newly enrolled clinicians were randomized by the last author (M.D.P.) to one of two supervision conditions stratified by supervisor (random number generator) to ensure condition balance within supervisors (see Figure 1). Study staff notified clinicians of their condition. Supervisor adherence to condition was monitored. Research assistants masked to condition reviewed four randomly selected supervision meetings per supervisor (two per condition, distributed across clinicians) within 2-weeks of the supervision meeting occurring. Experimental drift was exceedingly rare (fewer than five instances a year); however, when identified, supervisors were contacted within 2 weeks for booster training on conditions. Supervisors received an email on condition adherence every 2 weeks.

Intervention. The delivery of TF-CBT was the focus of supervision. TF-CBT is an EBT designed to address symptoms of posttraumatic stress (PTS), depression and disruptive behaviors among youth exposed to traumatic events (Cohen et al., 2017; Dorsey, McLaughlin et al., 2017). The acronym "PRACTICE" summarizes components clustered in three phases: (a) *stabilization* consisting of psychoeducation, parenting (for guardians), relaxation, affective modulation, cognitive coping (i.e., PRAC), (b) *trauma narration (TN) and processing* consisting of the initiation of discussion and/or writing about the traumatic event/s and cognitive processing (i.e., T), and (c) *integration and consolidation* including in-vivo mastery, conjoint caregiver–child sessions, and enhancing safety and future development (i.e., ICE). For our examination of adherence, we focused on stabilization, gradual exposure in stabilization, and TN (within the TN and processing phase) as well as days until TN initiation.

Description of Supervision Conditions

SAU/Phase I (Baseline). Clinicians who participated in Phase I of STEPS received SAU. Supervisors were asked to supervise clinicians' TF-CBT cases as they naturally would. Having a baseline supervision phase allows for examination of clinicians' TF-CBT delivery under usual supervision—without intervention—prior to introduction of the two RCT conditions in Phase II of STEPS. In the SAU phase, supervisors most often used techniques that included supportive listening, information gathering, didactic instruction, and providing clinical suggestions (Dorsey et al., 2018).

SFM. In the Phase II RCT, clinicians in both conditions monitored key symptoms in each session using brief measures standard in the CBT+-Initiative. These include the child-completed Screen for Child Anxiety-Related Emotional Disorders—Anxiety and PTS Subscales (Birmaher et al., 1997; Muris et al., 2000) and the caregiver-completed externalizing subscale of the Pediatric Symptom Checklist-17 (PSC-17) (Gardner et al., 2007). These measures were available through the deidentified, web-based measurement feedback system, EBP toolkit (www.ebptoolkit.com), used by all CBT+ training participants, where clinicians use monikers for their clients (e.g., skateboarder). We added a feature that automatically graphed measure scores over time, providing clinicians and supervisors with a pictorial representation of symptom improvement, maintenance, or deterioration. *Fidelity monitoring* was also conducted within EBP toolkit, via a short TF-CBT adherence checklist completed by clinicians after treatment sessions. Supervisors reviewed symptom and adherence data by accessing EBP toolkit during supervision. Supervisors had a slightly expanded TF-CBT checklist that included follow-up adherence queries, including key elements of the TF-CBT components and cross-cutting CBT techniques (e.g., homework, role-play), as well as a prompt on preparing for the upcoming session.

SFM + BR. In the Phase II RCT, clinicians in the SFM + BR condition received SFM and an additional, multi-purpose strategy—behavioral rehearsal (BR). The inclusion of BR is potentially a high-yield addition to address concerns of relying on clinician self-report of fidelity (i.e., indirect methods) and promote skill building (Bearman et al., 2013). In the SFM + BR condition, in each supervision session, supervisors had the clinician do a short BR (5–10 min) demonstrating how they would deliver a TF-CBT element relevant to an upcoming session, followed by supervisor feedback. A set of brief BR guidelines (e.g., bulleted form; one-page; see our Supplemental File) were provided to supervisors for TF-CBT components, along with pointers for addressing common challenges (e.g., dealing with avoidance, homework assignment/review).

Participants

Supervisors

Fifty-six supervisors were enrolled across Phases I and II; 48 of these supervised clinicians with recorded sessions

Figure 1
STEPS Consort Diagram
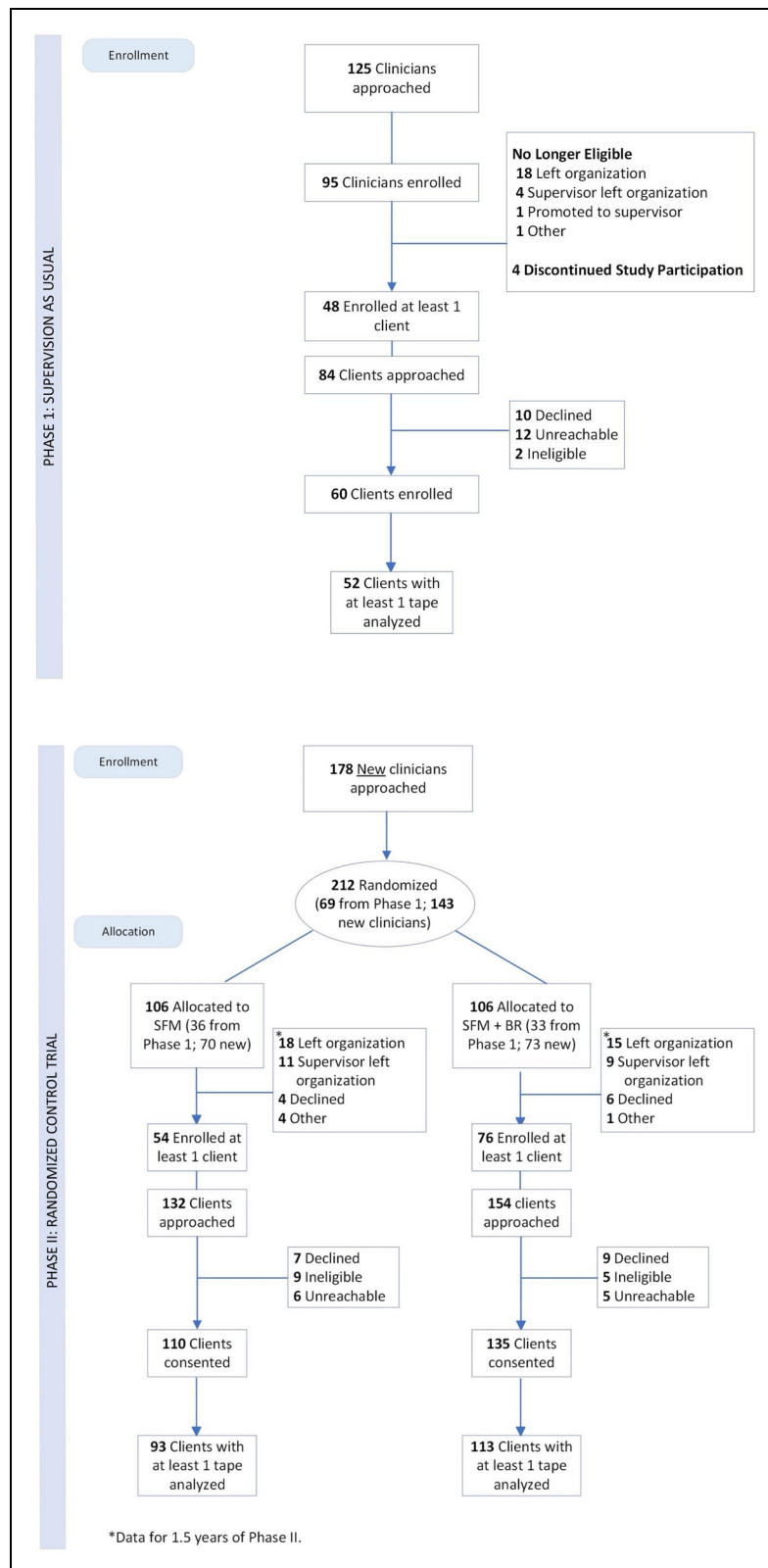
Note. STEPS = Supervision to Enhance Practice Study.

Table 1
Participant Characteristics[a]

	Clinicians (N = 151)	Supervisors (N = 48)	Youth (N = 305)
Race/ethnicity, n (%)			
White (non-Hispanic)	125 (82.8)	43 (89.6)	154 (50.5)
Asian	3 (2.0)	2 (4.2)	0
Black or African American	4 (2.7)	0	11 (3.6)
American Indian or Alaska native	0	0	6 (2.0)
Native Hawaiian or other Pacific Islander	1 (0.7)	1 (2.1)	0
Latinae/a/x or Hispanic	8 (5.3)	1 (2.1)	57 (18.7)
Multiracial	8 (5.3)	1 (2.1)	76 (24.9)
Other	2 (1.3)	0	1 (0.3)
[b]**Household income, n (%)**			
<$25,000	-	-	168 (55.1)
$25,000–$49,999	-	-	83 (27.2)
$50,000–$74,999	-	-	18 (5.9)
$75,000+	-	-	33 (10.8)
Medicaid recipient (%)	-	-	85.9
Female (%)	86.1	75.0	68.2
Age, M (SD)	35.9 (10.4)	41.2 (9.7)	11.7 (3.0)
Master's degree training (%)	93.4	93.8	-
Years providing psychotherapy, M (SD)	4.7 (5.2)	12 (7.1)	-
Years at organization, M (SD)	3.2 (3.4)	8.5 (6.2)	-
TF-CBT self-efficacy, M (SD)	3.09 (0.7)	-	-
TF-CBT objective knowledge, M (SD)	9.21 (2.0)	10.54 (1.5)	-
Self-efficacy in supervising TF-CBT, M (SD)	-	3.54 (0.5)	-

Note. TF-CBT = trauma-focused cognitive behavioral therapy.
TF-CBT self-efficacy and self-efficacy supervising TF-CBT scores range 1–5.
TF-CBT objective knowledge scores range 1–13.
[a]Participants in our analysis sample for TF-CBT fidelity outcomes.
[b]Household income information collected between 2013 and 2016.

(analysis sample). Table 1 provides demographic information for all participants. Supervisor participants were located in 29 CMHCs in 39 offices across WA. Of 70 supervisors approached, 78% were eligible and agreed to participate. Criteria for study inclusion were receiving TF-CBT-specific training as part of the EBT initiative and supervising two or more clinicians eligible to participate. There were no exclusionary criteria. Thirty-three supervisors were enrolled in Phase I. For Phase II, 25 of these continued and 23 additional supervisors were enrolled.

Clinicians

Two hundred thirty-eight clinicians were enrolled across phases I and II; 151 clinicians had some child data (clinical outcome measures), 133 of these had recorded TF-CBT sessions and are part of the analysis sample (see Figure 1). Ninety-five clinicians were enrolled in Phase I. For Phase II (RCT), 69 of these continued and an additional 143 clinicians were enrolled. Across both phases, of the 303 approached, 78% were eligible and agreed to participate. Criteria for study inclusion were receiving supervision from a participating supervisor and being trained in TF-CBT. Training was defined as: (a) receiving training through the initiative or (b) through completion of the freely available, 10-hr, online TF-CBT program (https://tfcbt.musc.edu/) and previous provision of TF-CBT to one client under supervision of a TF-CBT-trained supervisor at their organization. In-person, initiative training was not required for clinicians because annual training spaces are typically limited to two to three clinicians per organization. Given high turnover rates, organizations regularly rely on online training before clinicians can do in-person training. Exclusionary criteria were having an adult-only caseload, employed less than 80% time, and immediate plans to leave the organization.

Children and Adolescents

Participants were 305 youth and a guardian; 258 youth had recorded TF-CBT sessions and are part of the analysis sample. The study team contacted 370 families across both phases; 82.4% were eligible and agreed to participate. Reasons for nonparticipation and rates are included in Figure 1. Inclusion criteria were: (a) client of a participating clinician and receiving TF-CBT; (b) 6–17 age range; (c) endorsed trauma history; (d) significant baseline PTS symptoms assessed by the study team; (e) living with a parent/legal guardian willing to participate; and (f) spoke English (guardians could speak Spanish or English). Exclusion criteria were: (a) youth having a pervasive developmental disorder or substantial cognitive impairment and (b) guardian serious mental illness.

Measures

All supervisor and clinician self-report measures were administered upon participant enrollment in the study, and before engaging in any other study activities (e.g., supervision/client session recording, participation in RCT conditions). For most participants, enrollment was at the beginning of Phase I or II. However, given high rates of supervisor and clinician turnover, we continued to enroll participants during both study phases.

Participant Characteristics

Participants provided information on demographics, work environment, clinical experience, training, and other background information (Table 1). Supervisory-

specific information was also obtained (e.g., number of supervisees).

TF-CBT Knowledge

Supervisors and clinicians completed the 13-item TF-CBT knowledge test. Adapted from the Denver Post Health Survey (Fitzgerald, 2010), this measure reflects domains in the TF-CBT certification program. Participants responded to multiple choice and true–false items. In our sample, average item difficulty was 0.73 (range 0.32–0.93) and average item discrimination was 0.38 (range 0.20–0.53), showing a good relation between probability of answering individual items correctly and respondents' total score. Our measure also demonstrated good convergent validity with clinician-reported TF-CBT training extensiveness ($r = .42$, $p < .001$) and the related, but distinct construct of TF-CBT self-efficacy ($r = .27$, $p < .001$).

Self-efficacy in Delivering TF-CBT

Clinicians were asked to complete an 11-item index to assess their self-efficacy in TF-CBT (Deblinger & CARES Institute, 2013; NCVRTC MUSC, 2010). Participants rated perceived competency in implementing TF-CBT on a 5-point Likert scale (ranging from $0 = $ *not at all* to $4 = $ *exceptionally*). Participants rated their self-efficacy in TF-CBT skills such as, "Completing trauma narratives with children." Higher scores indicate greater self-efficacy in TF-CBT delivery. Cronbach's alpha for this sample was .92.

Self-efficacy in Supervising TF-CBT

Supervisors' perceived competence in supervising TF-CBT was measured using a 13-item self-efficacy in supervision index adapted from a measure created by TF-CBT developers (Deblinger & CARES Institute, R. U., 2013; NCVRTC, MUSC, 2010). Items were measured on a 5-point Likert scale ($1 = $ *not at all* to $5 = $ *exceptionally*). Sample items include an assessment of confidence to: "Supervise clinicians in all of the TF-CBT components." Higher scores indicate greater self-efficacy in TF-CBT supervision. Cronbach's alpha for this sample was .93.

PTS Symptoms

The 20-item UCLA Post Traumatic Stress Disorder Reaction Index (PTSD-RI; Steinberg et al., 2004) was administered by study staff to youth (ages 8 and higher) and their guardians to assess baseline PTS symptoms and determine youth study eligibility (severity cutoff of ≥ 21 or algorithm scoring for likelihood to meet diagnostic criteria). The UCLA PTSD-RI demonstrates good convergent validity and good to excellent test–retest reliability, with Cronbach's α .90 for the total scale (Steinberg et al., 2013).

Days until TN Receipt (Outcome Variable)

Days until TN receipt were calculated as days between the client's first session and the TF-CBT session focused on the TN component.

TF-CBT Adherence (Outcome Variables)

The 25-item, TF-CBT-specific version of the Therapeutic Process Observational Coding System for Child Psychotherapy (TPOCS-S Garland et al., 2010; McLeod & Weisz, 2010) was used to assess TF-CBT adherence by coding audiorecorded TF-CBT sessions. The measure includes 10 TF-CBT content area items (e.g., relaxation, TN), two general content items (i.e., assessment, other topics/crisis or case management), and 13 therapeutic techniques (e.g., assign/review homework, Socratic questioning). This study uses only the 10 TF-CBT content items. The scoring strategy includes extensiveness ratings to assess the degree to which clinicians use each element in a session, using a 7-point Likert-type extensiveness scale 0–6[1] ($0 = $ *not used* to $6 = $ *extensively*). To determine extensiveness, coders note element delivery in 5-min intervals using symbols to denote thoroughness within that interval. At the end of the session, coders assign a total extensiveness score by item that reflects frequency and thoroughness; thus extensiveness ratings provide dosage information about each TF-CBT component. Coders were masked to clinician/youth condition.

Coder Training/Therapy Session Sampling and Reliability

Coder Training. Nine postbaccalaureate research assistants were coders. They participated in a 2-day TF-CBT clinical training, completed a 10-hr TF-CBT online course, and read the TF-CBT treatment manual (Cohen et al., 2017). Then they received a multiday coder training, led by a TF-CBT treatment developer, an expert trainer (and study PI), and an experienced coder in prior TF-CBT studies. Coder training comprised of independent review of the coding manual, didactics, independent coding, group review of coded sessions, and joint listening to sessions to achieve consensus. For establishing coder readiness to code independently (i.e., acceptable interrater reliability across group members and with expert trainers), coding training leads independently coded 10 training files, reaching consensus on final codes. Trainee coders independently coded those 10 training files. Once ratings met the threshold for interrater reliability, ICC $(2,1) \geq .80$ (Cicchetti, 1994), coders began coding independently. If an individual item had an ICC $(2,1) \leq .60$, coders were assigned additional sessions for review and practice. To support high reliability over time, coders reviewed the coding manual monthly, attended periodic booster training, and received feedback on their reliability.

Interrater Reliability. Audiorecorded sessions were randomly assigned to coders. A subset of recordings (20%) were coded by two or more coders across multiple waves over time, using absolute consensus, single measures, and two-way random effects models. ICCs (2,2) were ≥ 0.70

for each wave and averaged .834 (ICC 2,9) indicating excellent reliability (Cicchetti, 1994).

Session Sampling Procedures. In total, 2,135 recordings of TF-CBT sessions were received from 151 clinicians and 305 unique youth clients (258 of whom had at least one recorded session). Each clinician submitted TF-CBT sessions for an average of 2.00 ($SD = 1.24$) clients. Across phases I and II, clinicians submitted an average of 8.02 recordings ($SD = 4.21$; range $= 1$–22). There was an average of 8.27 recordings submitted for each youth ($SD = 4.82$, range $= 1$–29), from which two were randomly selected. One focused on early treatment—stabilization phase—and one focused on later treatment and imaginal exposure, the TN phase. We had recorded stabilization sessions for 254 youth and recorded TN phase sessions for 160 youth (156 youth had stabilization and TN recordings). Stabilization phase extensiveness was calculated using the highest extensiveness score for any stabilization phase content item. TN phase extensiveness and occurrence were determined using the audiorecording labels submitted by clinicians. If a clinician never submitted an audio file labeled "trauma narrative" or no other session was deemed to include the TN based on coder review, they were considered to have not delivered this component.

Analysis Plan

Our analysis combined data from STEPS Phases I and II. Phase I data represents the SAU condition. Phase II included some continuing supervisors and clinicians from Phase I, and new supervisors and their clinicians. Therefore, some supervisors and clinicians contribute data to phases I and II. Descriptive statistics summarized participant demographics and baseline measures. To evaluate the effects of supervision conditions on adherence, we planned to use hierarchical generalized linear models to account for three-level clustering (i.e., client-level data within each clinician within their supervisor). However, the three-level model failed to converge in some analyses. For models that converged, the ICC of the null model showed that <5% of the total variance in outcomes occurred at the supervisor-level. We instead employed generalized estimating equations (GEEs), which allow for two-level clustering predicting adherence, using individual-mean fidelity scores for each client per clinician. Supervision condition was the main predictor, and SAU was the reference category.

Main Effects

Six GEEs were conducted to test the effect of supervision condition and clinician- and supervisor-level variables on outcomes of interest: (a) stabilization phase (i.e., PRAC) extensiveness, (b) gradual exposure occurrence during stabilization phase, (c) gradual exposure extensiveness during stabilization phase, (d) TN occurrence, (e) TN extensiveness, and (f) days until TN receipt. Stabilization phase occurrence was not included as an outcome because of lack of variability: the stabilization phase of TF-CBT involves the first set of treatment components and thus occurred for all cases that began treatment. When the outcome variable reflected occurrence (i.e., a binary yes/no outcome), we used binomial models; Gaussian models were used for *extensiveness* outcome variables. For each analysis, we included supervision condition and the following covariates: child baseline PTS, clinician TF-CBT objective knowledge, clinician TF-CBT self-efficacy, supervisor TF-CBT objective knowledge, and supervisor self-efficacy in supervising TF-CBT. For the model predicting TN occurrence, we added one additional covariate, the total number of sessions submitted by each clinician, as more sessions submitted means that clinicians likely delivered more TF-CBT sessions, which may naturally increase the likelihood of TN occurrence. Analyses were conducted in R 4.1.2 using the *gee* package. G*Power 3.1 was used to assess whether our analyses were powered for main effects and moderation analyses (Faul et al., 2009). Assuming power of 0.8, our sample size was powered to detect small-to-moderate effect sizes (Cohen's $f^2 = 0.1$). Given that our data had limited missingness ($\leq$5%), we used a listwise deletion approach.

Interaction Effects

We ran separate models to examine whether four covariates—clinician TF-CBT objective knowledge, clinician TF-CBT self-efficacy, supervisor TF-CBT objective knowledge, and supervisor self-efficacy in supervising TF-CBT—were potential moderators of supervision condition on our six outcomes (i.e., four potential moderators for six outcomes; total of 24 additional analyses). Each analysis had one interaction term (Potential Moderator × Supervision Condition). Similar to the main effects models, the only analyses with one additional covariate (i.e., total number of sessions submitted) were the ones where TN occurrence was the outcome. We used corrections to adjust for multiple testing to reduce false discovery rate (Benjamini & Hochberg, 1995).

Results

Randomization tests were conducted to ensure that the distribution of our clinician and child client variables were balanced across randomized conditions. There were no statistically significant differences across conditions ($p > .05$; Cohen's d range $= 0.06$–0.31). Table 1 presents participants' demographic information. Supervisors and clinicians were predominantly white (supervisors $= 90$%, clinicians $= 83$%), female (supervisors $= 75$%, clinicians $= 86$%), held a master's degree (supervisors $= 94$%, clinicians $= 93$%), and ascribed to a primarily CBT orientation (supervisors $= 73$%, clinicians $= 66$%). Youth were racially and ethnically diverse, mostly female, from households with an income less than $25,000, and receiving Medicaid-funded services. The average age for youth was 11.7 ($SD = 3.0$) and the average baseline PTS score was 36.8 ($SD = 14.2$).

Table 2

Summarizing Adherence Outcomes by Supervision Conditions

	SP extensiveness	[a]GE occurrence	[a]GE extensiveness	TN phase occurrence	TN phase extensiveness	Days until TN
	$n = 254$ youth M (SD)	$n = 254$ youth (%)	$n = 254$ youth M (SD)	$n = 160$ youth (%)	$n = 160$ youth M (SD)	$n = 160$ youth M (SD)
SAU	3.94 (1.36)	71	2.75 (1.27)	64	4.24 (0.79)	67.52 (43.91)
SFM	3.95 (1.28)	73	2.79 (1.31)	62	4.42 (1.17)	70.62 (34.93)
SFM + BR	4.02 (1.17)	71	2.68 (1.31)	61	4.35 (0.94)	64.10 (40.96)

Note. SP = stabilization phase components (PRAC components; from the TF-CBT PRACTICE acronym); GE = gradual exposure; TN = trauma narrative; TF-CBT = trauma-focused cognitive behavioral therapy; SFM + BR = symptom and fidelity monitoring plus behavioral rehearsal; SAU = supervision-as-usual.
Occurrence outcomes are frequencies and extensiveness outcomes are in means (range 1–6).
[a]GE adherence assessed during during the stabilization phase of TF-CBT.

On average, across supervision conditions, clinicians submitted around the same number of TF-CBT sessions ($M_{SAU} = 8.60$, $M_{SFM\ alone} = 8.32$, and $M_{SFM + BR} = 8.09$), and adherence outcomes were very similar (Table 2). For example, clinicians use of TN with youth was 64% in the SAU condition, 62% in TN in the SFM condition, and 61% in the SFM + BR condition, suggesting limited variation in TN occurrence across conditions. Moreover, days until TN receipt were also similar. The only adherence outcome with some (nonsignificant) variation by condition was TN extensiveness. TN extensiveness, and all other adherence scores, ranged from 0 to 6 in our data. The average level of TN extensiveness was 4.24 ($SD = 0.79$) in the SAU condition, 4.42 ($SD = 1.17$) in the SFM condition, and 4.35 ($SD = 0.94$) in the SFM + BR condition. In the SFM and SFM + BR conditions, clinicians had a .2 or .1 increase in TN extensiveness compared to SAU, but differences in TN extensiveness were not significant, $F(2,150) = 1.03$, $p = .36$. One-way ANOVAs were computed to determine effect sizes across conditions and all effect sizes were small (Cohen's f range = 0.03–0.07). Table 2 summarizes all fidelity outcomes by supervision condition.

Table 3A and B presents the estimates from the six GEE models examining effect of the supervision conditions (SFM and SFM + BR) relative to supervision-as-usual on TF-CBT adherence, controlling for hypothesized covariates. Two models had occurrence outcomes (i.e., binary): gradual exposure and TN phase *occurrence*. Adjusting for covariates, supervision condition was not significantly associated with either of the binary fidelity outcomes (Table 3A). The remainder of the GEE models had continuous outcome variables: stabilization phase extensiveness, gradual exposure extensiveness, TN extensiveness, and days till TN. Adjusting for covariates, supervision condition was not significantly associated with any of the continuous fidelity outcomes (Table 3B).

After adjusting for multiple comparisons, only one statistically significant interaction effect was found. Clinicians' TF-CBT self-efficacy moderated the relationship between supervision condition and trauma narration extensiveness. Specifically, higher clinician TF-CBT self-efficacy increased the effectiveness of SFM (vs. SAU) on TN extensiveness by 0.64 ($p = .03$) and of SFM + BR (vs. SAU) on TN extensiveness by 0.53 ($p = .04$; see Figure 2). None of the remaining moderators examined were significant (i.e., clinicians' TF-CBT knowledge or self-efficacy; supervisors' TF-CBT knowledge or self-efficacy in supervising).

Discussion

This is one of the first RCTs testing specific supervision techniques, delivered by workplace-based supervisors in community settings. We expected to find that systematic use of more active, gold-standard supervision techniques, such as those in our experimental conditions, would improve clinician treatment adherence. In particular, we expected that adding BR to SFM would have the greatest impact on clinician fidelity, given that this technique provides an opportunity for supervisors to observe how clinicians might deliver a particular treatment element to a client. However, our hypotheses were not supported; clinicians randomized to Phase II experimental conditions, SFM or SFM + BR, did not have higher adherence scores than those in the SAU phase, where adherence was already acceptable or high. Additionally, the SFM + BR condition did not outperform SFM. Only one of the moderators examined, clinician TF-CBT self-efficacy, clarified the relation between condition and adherence.

While these findings were unexpected, there may be a few reasons the experimental conditions did not outperform SAU. The most plausible explanation among these is that adherence scores on our six indicators in the SAU phase were already high. For example, the TN was delivered about 60% of the time. While that rate may be low for an efficacy trial, for CHMCs, where there can be high no-show rates and cancelations, and exposure use can be inconsistent (Borntrager et al., 2013; McLeod & Weisz, 2010), TN occurrence for 60% of the cases within

Table 3

(A) Main Effects Model Estimates for Relations Between Covariates and Adherence (Binary Outcomes)

	Occurrence outcomes variables: OR (SE)	
	Gradual exposure in stabilization phase $n = 254$ youth	Trauma Narrative phase $n = 160$ youth
Number of tapes submitted	N/A	[a]1.47 (0.06)
Baseline PTS severity score	1.01 (0.1)	0.98 (0.01)
SFM + BR	1.24 (0.41)	1.39 (0.46)
SFM	1.15 (0.42)	1.26 (0.48)
Clinician TF-CBT knowledge	0.99 (0.08)	0.92 (0.09)
Clinician TF-CBT self-efficacy	1.23 (0.24)	1.33 (0.27)
Supervisor TF-CBT knowledge	0.93 (0.12)	1.01 (0.12)
Supervisor self-efficacy in supervising TF-CBT	1.42 (0.40)	0.98 (0.40)

Note. OR = odds ratio; SE = standard error; TF-CBT = trauma-focused cognitive behavioral therapy; PTS = posttraumatic stress; SFM + BR = symptom and fidelity monitoring plus behavioral rehearsal; SAU = supervision-as-usual.
For conditions SFM and SFM + BR, the reference category was SAU.
[a]Estimate indicates statistical significance ($p < .001$).

Table 3

(Continued) (B) Main Effects Model Estimates for the Relation Between Covariates and Adherence (Continuous Outcomes)

	Extensiveness outcome variables: Beta (SE)			
	Stabilization phase $n = 254$ youth	Gradual exposure in stabilization phase $n = 254$ youth	Trauma Narrative phase $n = 160$ youth	Days until TN $n = 160$ youth
Baseline PTS severity score	−0.01 (0.01)	0.004 (0.01)	−0.004 (0.01)	−0.03 (0.25)
SFM + BR	0.16 (0.23)	−0.04 (0.26)	0.15 (0.23)	0.901 (0.15)
SFM	0.09 (0.23)	0.12 (0.28)	0.35 (0.25)	5.03 (9.55)
Clinician TF-CBT knowledge	−0.01 (0.01)	0.08 (0.05)	−0.03 (0.04)	−4.43 (1.76)
Clinician TF-CBT self-efficacy	0.21 (0.13)	0.02 (0.15)	[a]0.43 (0.11)	−4.45 (5.05)
Supervisor TF-CBT knowledge	0.03 (0.24)	−0.05 (0.07)	−0.01 (0.08)	5.62 (3.04)
Supervisor self-efficacy in supervising TF-CBT	0.19 (0.19)	0.10 (0.23)	−0.22 (0.20)	−7.00 (8.62)

Note. SE = standard error; TN = trauma narrative; TF-CBT = trauma-focused cognitive behavioral therapy; SFM + BR = symptom and fidelity monitoring plus behavioral
rehearsal; SAU = supervision-as-usual.
For conditions SFM and SFM + BR, the reference category was SAU.
[a]Estimate indicates statistical significance ($p < .001$).

6 months is positive. Perhaps most surprisingly, the *dosage or extensiveness* (McLeod et al., 2015) of TN delivery during the SAU was also quite high. Considering a common threshold of 70% as indicative of acceptable levels of fidelity where treatment may be potent enough to be beneficial for clients (Dorsey, Lyon et al., 2017); the TN extensiveness scores in the SAU phase (4.24 on our 0–6 scale, ≥70% of the total possible score) meet this threshold. In two studies using similar extensiveness coding systems to examine child anxiety treatment adherence in research settings (efficacy trials) versus community settings, our rates were closer to the efficacy trial ratings (McLeod et al., 2019; Smith et al., 2017).

What might have supported this high rate of adherence in SAU? We suspect the context of our longstanding state-funded EBT initiative offers some explanation. While the funding is modest (200–250k), it supports EBT expertise through ongoing training and consultation for clinicians, with expectations for documented CBT delivery (in the online, deidentified dashboard system). The initiative also explicitly supports supervisors, offering supervisor-specific learning and training opportunities. Supervisors in CMHCs help lead the initiative and provide training. If the initiative supports are partially responsible for the high SAU adherence, our findings provide some good news for what is possible for EBT in other states and systems. However, without financial support, these activities and potential use of the supervision strategies from our experimental conditions could be challenging to implement. The structure of TF-CBT may also have supported

Figure 2

Clinician TF-CBT Self-efficacy Moderates the Effect of Supervision Condition on Trauma Narrative Extensiveness
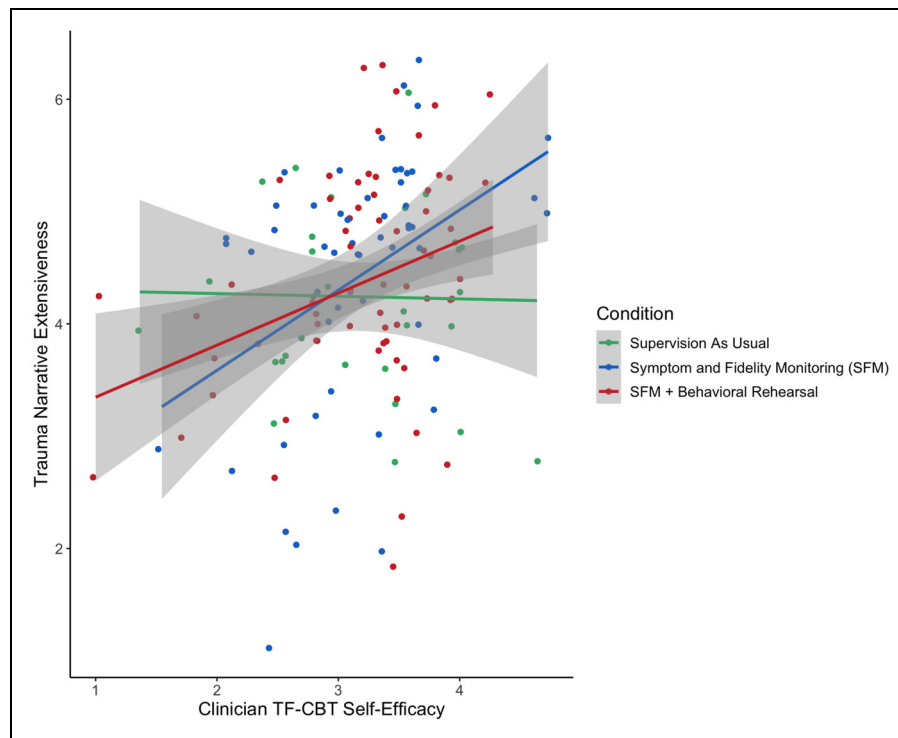
Note. TF-CBT = trauma-focused cognitive behavioral therapy.

clinicians' high levels of adherence in SAU, given clearly defined elements and ways to conduct exposure with clients. It is also likely that the supervisors who enrolled in the study were highly motivated to provide adherence-focused supervision, and/or that clinicians' awareness of the recording of sessions and participation in research may have influenced the high adherence rates we saw in our study.

Another plausible explanation for nonsignificant differences across conditions, given other analyses on SAU is that examining the dose of the supervision techniques and/or different clusters of techniques used in conjunction might be a more nuanced approach. In Meza et al.'s (2023) examination of naturally occurring clusters, supervision that included frequent use of directive techniques (e.g., clinical suggestions, didactic instruction), was associated with higher TN delivery. A third explanation is that there may be important moderators of experimental condition effects that we did not examine (e.g., clinician motivation, TF-CBT buy-in, or burnout). However, this seems unlikely given that among moderators we did examine, we found evidence for only one: clinician self-efficacy and TN extensiveness, clinicians reporting higher self-efficacy in either experimental condition—SFM or SFM+BR—demonstrated higher TN extensiveness compared to SAU. Potentially, this suggests that clinicians who perceive themselves as more capable in TF-CBT may be more "ready" to take advantage of active supervision techniques like SFM and/or BR. There was no support for any of the supervisor-level moderators.

Limitations

The study has limitations that should be considered. First, clinician attrition in CMHCs can be high (30–60%/year; Barak et al., 2001; Beidas et al., 2016). Throughout our study, a substantial number of clinicians left their organizations and thus we had fewer youth per clinician than planned (as clinicians who left had shorter periods to enroll youth). We cannot be sure that enrolled clients are representative of clinicians' practice with other clients who received TF-CBT. We also cannot be sure that clinicians and supervisors did not give more attention to TF-CBT with study-enrolled clients. While youth participant diversity was high, as in many CMHCs, supervisors, and clinicians were mostly White. Although CBT+ state support is modest, many states do not have initiatives or the developed EBT expertise to rely on workplace-based supervisors. Our findings would be most generalizable to community mental health settings where supervision occurs regularly (e.g., on a weekly basis) and providers have received some training in EBTs. Our

findings would be least generalizable to settings where supervision is infrequent or non-existent. Treatment adherence is an important outcome; however, we did not examine clinician competence (were BR could be most beneficial) or clinical outcomes. The true "test" of any implementation strategy is if outcomes for clients are improved. Finally, we note that we did not control for baseline adherence levels for clinicians who participated in Phase I SAU and the Phase II RCT.

Conclusions

Clinicians in publicly funded CMHCs who participated in this study appeared to deliver TF-CBT well, within the context of our supportive state-funded EBT initiative. In other contexts or for other EBTs where there is not an established EBT initiative with similar supports, systematic use of techniques in our experimental supervision conditions may have an effect on adherence, or other outcomes like competence or clinical outcomes. Within our sample, we did not see an effect of experimental supervision conditions examined, in large part due to clinicians' high-quality delivery in the supervision-as-usual phase. If this is mostly due to the state-funded EBT initiative with its training method and focus on building and supporting strong supervisors, this would actually be good news for usual care, academic–community partnerships and implementation supports that may be needed to support EBT. We imagine aspects of our state-supported EBT initiative would be highly relevant for supporting other EBTs.

Author's Note

Rashed AlRasheed is also affiliated at Department of Community Medicine and Behavioral Sciences, College of Medicine, Kuwait University, Kuwait City, Kuwait.

Noah Triplett is also affiliated at Department of Psychology, University of Maryland, College Park, MD, USA.

ORCID iDs

Shannon Dorsey https://orcid.org/0000-0001-6240-0294
Suzanne EU Kerns https://orcid.org/0000-0003-4462-6817
Noah Triplett https://orcid.org/0000-0003-2325-3599
Michael D Pullmann https://orcid.org/0000-0001-8404-864X

Supplemental Material

Supplemental material for this article is available online.

Note

1. The TPOCS-S also uses a 7-point likert-type scale, but from 1 to 7. To compare scores with other TPOCS-S measures, the scale difference (0–6 vs. 1–7) should be considered.

References

Ashwood, J. S., Kataoka, S. H., Eberhart, N. K., Bromley, E., Zima, B. T., Baseman, L., Marti, F. A., Kofner, A., Tang, L., Azhar, G. S., Chamberlin, M., Erickson, B., Choi, K., Zhang, L., Miranda, J., & Burnam, M. A. (2018). Evaluation of the Mental Health Services Act in Los Angeles county: Implementation and outcomes for key programs. *Rand Health Quarterly*, *8*(1). https://doi.org/10.7249/rr2327

Bambling, M., King, R., Raue, P., Schweitzer, R., & Lambert, W. (2006). Clinical supervision: Its influence on client-rated working alliance and client symptom reduction in the brief treatment of major depression. *Psychotherapy Research*, *16*(3), 317–331. https://doi.org/10.1080/10503300500268524

Bandura, A. (1998). Health promotion from the perspective of social cognitive theory. *Psychology and Health*, *13*(4), 623–649. https://doi.org/10.1080/08870449808407422

Barak, M. E. M., Nissly, J. A., & Levin, A. (2001). Antecedents to retention and turnover among child welfare, social work, and other human service employees: What can we learn from past research? A review and metaanalysis. *Social Service Review*, *75*(4), 625–661. https://doi.org/10.1086/323166

Bearman, S. K., Schneiderman, R. L., & Zoloth, E. (2017). Building an evidence base for effective supervision practices: An analogue experiment of supervision to increase EBT fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*(2), 293–307. https://doi.org/10.1007/S10488-016-0723-8

Bearman, S. K., Weisz, J. R., Chorpita, B. F., Hoagwood, K., Ward, A., Ugueto, A. M., & Bernstein, A. (2013). More practice, less preach? The role of supervision processes and therapist characteristics in EBP implementation. *Administration and Policy in Mental Health and Mental Health Services Research*, *40*(6), 518–529. https://doi.org/10.1007/S10488-013-0485-5

Beidas, R. S., Cross, W., & Dorsey, S. (2014). Show me, don't tell me: Behavioral rehearsal as a training and analogue fidelity tool. *Cognitive and Behavioral Practice*, *21*(1), 1–11. https://doi.org/10.1016/J.CBPRA.2013.04.002

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 (Washington, D.C.), *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

Beidas, R. S., Marcus, S., Wolk, C. B., Powell, B., Aarons, G. A., Evans, A. C., Hurford, M. O., Hadley, T., Adams, D. R., Walsh, L. M., Babbar, S., Barg, F., & Mandell, D. S. (2016). A prospective examination of clinician and supervisor turnover within the context of implementation of evidence-based practices in a publicly-funded mental health system. *Administration and Policy in Mental Health and Mental Health Services Research*, *43*(5), 640–649. https://doi.org/10.1007/S10488-015-0673-6

Beidas, R. S., Williams, N. J., Becker-Haimes, E. M., Aarons, G. A., Barg, F. K., Evans, A. C., Jackson, K., Jones, D., Hadley, T., Hoagwood, K., Marcus, S. C., Neimark, G., Rubin, R. M., Schoenwald, S. K., Adams, D. R., Walsh, L. M., Zentgraf, K., & Mandell, D. S. (2019). A repeated cross-sectional study of clinicians' use of psychotherapy techniques during 5 years of a system-wide effort to implement evidence-based practices in Philadelphia. *Implementation Science*, *14*(1), 1–13. https://doi.org/10.1186/S13012-019-0912-4/TABLES/3

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/J.2517-6161.1995.TB02031.X

Birmaher, B., Khetarpal, S., Brent, D., Cully, M., Balach, L., Kaufman, J., & Neer, S. M. K. (1997). The Screen for Child Anxiety Related Emotional Disorders (SCARED): Scale construction and psychometric characteristics. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*(4), 545–553. https://doi.org/10.1097/00004583-199704000-00018

Bitsko, R. H., Claussen, A. H., Lichstein, J., Black, L. I., Jones, S. E., Danielson, M. L., Hoenig, J. M., Davis Jack, S. P., Brody, D. J., Gyawali, S., Maenner, M. J., Warner, M., Holland, K. M., Perou, R., Crosby, A. E., Blumberg, S. J., Avenevoli, S., Kaminski, J. W., & Ghandour, R. M. (2022). Mental health surveillance among children - United States, 2013-2019. *MMWR Supplements*, *71*(2), 1–42. https://doi.org/10.15585/MMWR.SU7102A1

Borntrager, C., Chorpita, B. F., Higa-McMillan, C. K., Daleiden, E. L., & Starace, N. (2013). Usual care for trauma-exposed youth: Are clinician-reported therapy techniques evidence-based? *Children and Youth Services Review*, *35*(1), 133–141. https://doi.org/10.1016/j.childyouth.2012.09.018

Bruns, E. J., Kerns, S. E. U., Pullmann, M. D., Hensley, S. W., Lutterman, T., & Hoagwood, K. E. (2016). Research, data, and evidence-based treatment use in state behavioral health systems, 2001-2012. *Psychiatric Services (Washington, D.C.)*, *67*(5), 496–503. https://doi.org/10.1176/APPI.PS.201500014

Choy-Brown, M., & Stanhope, V. (2018). The availability of supervision in routine mental health care. *Clinical Social Work Journal*, *46*(4), 271–280. https://doi.org/10.1007/S10615-018-0687-0

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Cohen, J. A., Mannarino, A. P., & Deblinger, E. (2017). *Treating trauma and traumatic grief in children and adolescents* (2nd ed.). Guilford Press.

Cross, W., Matthieu, M. M., Cerel, J., & Knox, K. L. (2007). Proximate outcomes of gatekeeper training for suicide prevention in the workplace. *Suicide and Life-Threatening Behavior*, *37*(6), 659–670. https://doi.org/10.1521/suli.2007.37.issue-6

Deblinger, E., & CARES Institute, R. U. (2013). *Unpublished measure*.

Dorsey, S., Berliner, L., Lyon, A. R., Pullmann, M. D., & Murray, L. K. (2016). A statewide common elements initiative for children's mental health. *The Journal of Behavioral Health Services & Research*, *43*(2), 246–261. https://doi.org/10.1007/s11414-014-9430-y

Dorsey, S., Kerns, S. E. U., Lucid, L., Pullmann, M. D., Harrison, J. P., Berliner, L., Thompson, K., & Deblinger, E. (2018). Objective coding of content and techniques in workplace-based supervision of an EBT in public mental health. *Implementation Science*, *13*(1), 21. https://doi.org/10.1186/s13012-017-0708-3

Dorsey, S., Lyon, A. R., Pullmann, M. D., Jungbluth, N., Berliner, L., & Beidas, R. (2017). Behavioral rehearsal for analogue fidelity: Feasibility in a state-funded children's mental health initiative. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*(3), 395–404. https://doi.org/10.1007/S10488-016-0727-4

Dorsey, S., McLaughlin, K. A., Kerns, S. E. U., Harrison, J. P., Lambert, H. K., Briggs, E. C., Revillion Cox, J., & Amaya-Jackson, L. (2017). Evidence base update for psychosocial treatments for children and adolescents exposed to traumatic events. *Journal of Clinical Child and Adolescent Psychology: The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53*, *46*(3), 303–330. https://doi.org/10.1080/15374416.2016.1220309

Dorsey, S., Pullmann, M. D., Deblinger, E., Berliner, L., Kerns, S. E., Thompson, K., Unützer, J., Weisz, J. R., & Garland, A. F. (2013). Improving practice in community-based settings: A randomized trial of supervision – study protocol. *Implementation Science*, *8*(1), 1–11. https://doi.org/10.1186/1748-5908-8-89

Dorsey, S., Pullmann, M. D., Kerns, S. E. U., Jungbluth, N., Meza, R., Thompson, K., & Berliner, L. (2017). The juggling act of supervision in community mental health: Implications for supporting evidence-based treatment. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*(6), 838–852. https://doi.org/10.1007/S10488-017-0796-Z

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149/METRICS

Fitzgerald, M. M. (2010). *Unpublished measure*. Denver Post Health Survey.

Fukui, S., Wu, W., & Salyers, M. P. (2019). Impact of supervisory support on turnover intention: The mediating role of

burnout and job satisfaction in a longitudinal study. *Administration and Policy in Mental Health and Mental Health Services Research*, 46(4), 488–497. https://doi.org/10.1007/S10488-019-00927-0

Gardner, W., Lucas, A., Kolko, D. J., & Campo, J. V. (2007). Comparison of the PSC-17 and alternative mental health screens in an at-risk primary care sample. *Journal of the American Academy of Child and Adolescent Psychiatry*, 46(5), 611–618. https://doi.org/10.1097/CHI.0B013E318032384B

Garland, E. L., Fredrickson, B., Kring, A. M., Johnson, D. P., Meyer, P. S., & Penn, D. L. (2010). Upward spirals of positive emotions counter downward spirals of negativity: Insights from the broaden-and-build theory and affective neuroscience on the treatment of emotion dysfunctions and deficits in psychopathology. *Clinical Psychology Review*, 30(7), 849–864. https://doi.org/10.1016/J.CPR.2010.03.002

Glasgow, R. E. (2013). What does it mean to be pragmatic? Pragmatic methods, measures, and models to facilitate research translation. *Health Education & Behavior : The Official Publication of the Society for Public Health Education*, 40(3), 257–265. https://doi.org/10.1177/1090198113486805

Henggeler, S. W., Schoenwald, S. K., Liao, J. G., Letourneau, E. J., & Edwards, D. L. (2002). Transporting efficacious treatments to field settings: The link between supervisory practices and therapist fidelity in MST programs. *Journal of Clinical Child and Adolescent Psychology : The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53*, 31(2), 155–167. https://doi.org/10.1207/S15374424JCCP3102_02

Huijg, J. M., Crone, M. R., Verheijden, M. W., Van Der Zouwe, N., Middelkoop, B. J., & Gebhardt, W. A. (2013). Factors influencing the adoption, implementation, and continuation of physical activity interventions in primary health care: A Delphi study. *BMC Family Practice*, 14(1), 1–9. https://doi.org/10.1186/1471-2296-14-142/TABLES/1

Kendall, P. C., & Frank, H. E. (2018). Implementing evidence-based treatment protocols: Flexibility within fidelity. *Clinical Psychology : A Publication of the Division of Clinical Psychology of the American Psychological Association*, 25(4). https://doi.org/10.1111/CPSP.12271

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development* (Vol. 1). Prentice-Hall.

Martino, S., Paris, M., Añez, L., Nich, C., Canning-Ball, M., Hunkele, K., Olmstead, T. A., & Carroll, K. M. (2016). The effectiveness and cost of clinical supervision for motivational interviewing: A randomized controlled trial. *Journal of Substance Abuse Treatment*, 68, 11–23. https://doi.org/10.1016/J.JSAT.2016.04.005

McLeod, B. D., Cox, J. R., Jensen-Doss, A., Herschell, A., Ehrenreich-May, J., & Wood, J. J. (2018a). Proposing a mechanistic model of clinician training and consultation. *Clinical Psychology : A Publication of the Division of Clinical Psychology of the American Psychological Association*, 25(3), Article e12260. https://doi.org/10.1111/CPSP.12260

McLeod, B. D., Cox, J. R., Jensen-Doss, A., Herschell, A., Ehrenreich-May, J., & Wood, J. J. (2018b). Proposing a mechanistic model of clinician training and consultation. *Clinical Psychology : A Publication of the Division of Clinical*

Psychology of the American Psychological Association, 25(3). https://doi.org/10.1111/CPSP.12260

McLeod, B. D., Smith, M. M., Southam-Gerow, M. A., Weisz, J. R., & Kendall, P. C. (2015). Measuring treatment differentiation for implementation research: The therapy process observational coding system for child psychotherapy revised strategies scale. *Psychological Assessment*, 27(1), 314–325. https://doi.org/10.1037/PAS0000037

McLeod, B. D., Southam-Gerow, M. A., Jensen-Doss, A., Hogue, A., Kendall, P. C., & Weisz, J. R. (2019). Benchmarking treatment adherence and therapist competence in individual cognitive-behavioral treatment for youth anxiety disorders. *Journal of Clinical Child and Adolescent Psychology : The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53*, 48(Suppl 1), S234–S246. https://doi.org/10.1080/15374416.2017.1381914

McLeod, B. D., & Weisz, J. R. (2010). The therapy process observational coding system for child psychotherapy strategies scale. *Journal of Clinical Child and Adolescent Psychology*, 39(3), 436–443. https://doi.org/10.1080/15374411003691750

Meza, R. D., AlRasheed, R., Pullmann, M. D., & Dorsey, S. (2023). Clinical supervision approach predicts evidence-based trauma treatment delivery in children's mental health. *Frontiers in Psychiatry*, 13, Article 1072844. https://doi.org/10.3389/FPSYT.2022.1072844/BIBTEX

Miller, W. R., Yahne, C. E., Moyers, T. B., Martinez, J., & Pirritano, M. (2004). A randomized trial of methods to help clinicians learn motivational interviewing. *Journal of Consulting and Clinical Psychology*, 72(6), 1050–1062. https://doi.org/10.1037/0022-006X.72.6.1050

Milne, D. (2007). An empirical definition of clinical supervision. *British Journal of Clinical Psychology*, 46(Pt 4), 437–447. https://doi.org/10.1348/014466507X197415

Milne, D. (2009). *Evidence-based clinical supervision: Principles and practice*. Blackwell Publishing; British Psychological Society.

Milne, D., & James, I. (2000). A systematic review of effective cognitive-behavioural supervision. *British Journal of Clinical Psychology*, 39(2), 111–127. https://doi.org/10.1348/014466500163149

Milne, D., & Reiser, R. P. (2012). A rationale for evidence-based clinical supervision. *Journal of Contemporary Psychotherapy*, 42(3), 139–149. https://doi.org/10.1007/S10879-011-9199-8/METRICS

Muris, P., Meesters, C., Merckelbach, H., & Hülsenbeck, P. (2000). Worry in children is related to perceived parental rearing and attachment. *Behaviour Research and Therapy*, 38(5), 487–497. https://doi.org/10.1016/S0005-7967(99)00072-8

National Crime Victims Research and Treatment Center (NCVRTC) MUSC. (2010). *Unpublished measure*.

Pullmann, M. D., Lucid, L., Harrison, J. P., Martin, P., Deblinger, E., Benjamin, K. S., & Dorsey, S. (2018). Implementation climate and time predict intensity of supervision content related to evidence-based treatment. *Frontiers in Public Health*, 6, Article 280. https://doi.org/10.3389/FPUBH.2018.00280

Rakovshik, S. G., McManus, F., Vazquez-Montes, M., Muse, K., & Ougrin, D. (2016). Is supervision necessary? Examining the effects of internet-based CBT training

with and without supervision. *Journal of Consulting and Clinical Psychology*, *84*(3), 191–199. https://doi.org/10.1037/CCP0000079

Roth, A. D., Pilling, S., & Turner, J. (2010). Therapist training and supervision in clinical trials: Implications for clinical practice. *Behavioural and Cognitive Psychotherapy*, *38*(3), 291–302. https://doi.org/10.1017/S1352465810000068

Schoenwald, S. K., Chapman, J. E., Kelleher, K., Hoagwood, K. E., Landsverk, J., Stevens, J., Glisson, C., Rolls-Reutz, J., Weisz, J. R., Chorpita, B. E., Gibbons, R., Green, E. P., Hoagwood, K., Jensen, P. S., Mayberg, S., Miranda, J., & Palinkas, L. (2008). A survey of the infrastructure for children's mental health services: Implications for the implementation of empirically supported treatments (ESTs). *Administration and Policy in Mental Health and Mental Health Services Research*, *35*(1-2), 84–97. https://doi.org/10.1007/S10488-007-0147-6

Schoenwald, S. K., Sheidow, A. J., & Chapman, J. E. (2009). Clinical supervision in treatment transport: Effects on adherence and outcomes. *Journal of Consulting and Clinical Psychology*, *77*(3), 410–421. https://doi.org/10.1037/A0013788

Sholomskas, D. E., Syracuse-Siewert, G., Rounsaville, B. J., Ball, S. A., Nuro, K. F., & Carroll, K. M. (2005). We don't train in vain: A dissemination trial of three strategies of training clinicians in cognitive–behavioral therapy. *Journal of Consulting and Clinical Psychology*, *73*(1), 106–115. https://doi.org/10.1037/0022-006X.73.1.106

Smith, M. M., McLeod, B. D., Southam-Gerow, M. A., Jensen-Doss, A., Kendall, P. C., & Weisz, J. R. (2017). Does the delivery of CBT for youth anxiety differ across research and practice settings? *Behavior Therapy*, *48*(4), 501–516. https://doi.org/10.1016/j.beth.2016.07.004

Steinberg, A. M., Brymer, M. J., Decker, K. B., & Pynoos, R. S. (2004). The University of California at Los Angeles post-traumatic stress disorder reaction index. *Current Psychiatry Reports*, *6*(2), 96–100. https://doi.org/10.1007/S11920-004-0048-2

Steinberg, A. M., Brymer, M. J., Kim, S., Briggs, E. C., Ippen, C. G., Ostrowski, S. A., Gully, K. J., & Pynoos, R. S. (2013). Psychometric properties of the UCLA PTSD Reaction Index: Part I. *Journal of Traumatic Stress*, *26*(1), 1–9. https://doi.org/10.1002/JTS.21780

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

Tugendrajch, S. K., Cho, E., Andrews, J. H., & Hawley, K. M. (2023). Characterizing supervision-as-usual: Findings from two provider surveys. *Psychological Services*, *20*(2), 248–255. https://doi.org/10.1037/SER0000747

Whitney, D. G., & Peterson, M. D. (2019). US National and state-level prevalence of mental health disorders and disparities of mental health care use in children. *JAMA Pediatrics*, *173*(4), 389–391. https://doi.org/10.1001/JAMAPEDIATRICS.2018.5399