An Exploration of Workplace-Based Supervision Styles and their Relation to Clinician EBT Delivery

Rosemary D. Meza

A dissertation submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

University of Washington
2021

Reading Committee:
Shannon Dorsey, Chair
Wendy Stone
Michael Pullmann

Program Authorized to Offer Degree:
Department of Psychology

University of Washington

Abstract

An Exploration of Workplace-Based Supervision Styles and their Relation to Clinician EBT Delivery

Rosemary D. Meza

Chair of the Supervisory Committee:

Shannon Dorsey

Department of Psychology

In community mental health settings, clinical supervision is nearly ubiquitous, yet workplace-based supervision is understudied in the context of evidence-based treatment (EBT) implementation. Recent research has characterized various techniques used by supervisors to support clinicians' EBT delivery. However, little is known about how workplace-based supervisors use combinations of techniques to support clinicians' EBT delivery. This study aimed to 1) examine the presence of 'supervision styles' comprised of various supervision techniques, 2) examine factors that predict the use of 'supervision styles,' and 3) examine whether those 'supervision styles' predict clinicians' EBT delivery.

Data come from a descriptive study of supervision in the context of a state-funded EBT implementation effort. Participants were 28 supervisors, 70 clinician supervisees, and 60 youth and their guardians from 17 public mental health organizations. Supervisors and clinicians completed a baseline survey, supervisors audio recorded supervision sessions

over 1 year, and clinicians audio recorded Trauma-Focused Cognitive Behavioral Therapy (TF-CBT) treatment sessions with youth for six months. Audio recordings of 438 supervision sessions were objectively coded for presence and intensity of 13 supervision techniques. Audio recordings of 465 treatment sessions were objectively coded for TF-CBT fidelity. Agglomerative hierarchical cluster analysis was used to identify clusters of supervision techniques used in supervision. Mixed effects logistic regression models were used to examine predictors of supervision clusters. Finally, generalized estimating equations were used to examine the relation among supervision clusters and fidelity to TF-CBT, and logistic regression was used to examine the relation among supervision clusters and the delivery of the trauma narrative – the imaginal exposure element of TF-CBT.

Results revealed two supervision clusters termed "*directive*" supervision and "*undifferentiated low*" supervision. The odds of a supervision session being "*directive*" were higher when the supervision session was longer, involved discussion of fewer clients, and the clinician had less experience delivering TF-CBT in the past three months. Clinicians who received a higher proportion of "*directive*" supervision sessions had greater odds of delivering the trauma narrative with a client.

To our knowledge, this was the first study to examine the presence of subgroups of supervision based on objectively coded supervision techniques used to support clinician EBT delivery in community mental health. The results suggest that when clinicians have less EBT experience, supervisors are more likely to use a "*directive*" form of supervision, and this supervision is associated with delivery of imaginal exposure.

Table of Contents

CHAPTER 1

SUPERVISION STYLES AMONG WORKPLACE-BASED SUPERVISION OF EBT

Introduction

Clinical supervision is broadly defined as an intervention carried out by a senior clinician with a more junior clinician that is evaluative in nature, ongoing, and serves to enhance the junior clinician's professional competency and monitor the quality of services they provide (Milne, 2009). More specifically, it has been described as the formal process of professional support and learning which enables practitioners to acquire knowledge and competence, assume responsibility for their own practices, and enhance the protection and safety of consumer care in complex situations (Department of Health, 1993). Supervision is as a necessary support to positively impact providers' behavior after receiving training in evidence-based treatment (EBT; Beidas & Kendall, 2010; Herschell, Kolko, Baumann, & Davis, 2010). It is the foremost method by which mental health professionals receive training in therapeutic practices (Falender et al., 2004) and also serves as a form of treatment quality assurance (Perepletchikova, Treat, & Kazdin, 2007). Clinical supervision focused on EBT has been shown to improve clinicians' EBT attitudes, knowledge, skill, and fidelity (Bearman, Schneiderman, & Zoloth, 2016; Henggeler et al., 2002). Despite the importance of clinical supervision, research is lacking on the specific workplace-based supervision techniques that best support clinicians' delivery of EBT.

What We Know About Supervision Techniques from the Literature

A number of techniques can be gleaned from the literature on efficacy trials, supervision models, clinician training, and consultation. These techniques provide some guidance on supervision techniques that may be relevant in workplace-based supervision of EBT.

'Gold-Standard' Techniques from Efficacy Trials

Efficacy trials of clinical interventions typically use a common set of 'gold standard' supervision techniques to support clinicians' treatment delivery and to enhance client

outcomes (Roth, Pilling, & Turner, 2010; Padesky, 1996; Milne, 2008). Common techniques include skill-building/behavioral rehearsal, fidelity monitoring, live or recorded review of practice, and client symptom monitoring. Efficacy trials typically ensure adequate delivery of these techniques for all clinicians, precluding a determination of the effectiveness of these techniques in influencing clinicians' EBT provision.

Techniques Proposed in Supervision Models

Supervision models provide guidance on supervision techniques that may be applicable to workplace-based supervision of EBTs. Milne (2009) proposed that clinicians achieve competence through engaging in experiential learning. Specifically, he argues that clinical competence is acquired through a combination of four components of the experiential learning cycle defined in Kolb's (1984) theory of experiential learning: experiencing, reflecting, conceptualizing, and planning. Supervisors likely need to use a variety of methods to effectively aid clinicians to use these modes of experiential learning to improve their ability to provide treatment competently, safely, and effectively (Milne & James, 2000). To develop their model of supervision, Milne and colleagues (2008) reviewed the supervision literature using a best-evidence synthesis approach. They distilled best practices from 24 studies, with adequate rigor, that manipulated supervision and identified 26 supervision techniques. The most frequently occurring techniques were teaching and instruction (75% of studies), corrective feedback (63%), live or recorded observation (42%), goal setting (38%), and question-and-answer methods (e.g., information-gathering; clarifying: 38%).

Techniques Used in EBT Training

Literature examining training methods to improve clinician delivery of EBTs provides insights into practices that may also be leveraged in supervision to support EBT delivery. A review of studies examining training methods suggests that passive didactic methods of training have limited effects on behavior, and active learning strategies are required to

influence therapists behavior and client change (Beidas & Kendall, 2010). Active learning strategies are those that require the trainee to participate in the skill being taught, such as through modeling, behavioral rehearsals, and receiving coaching and feedback. Regardless of the strategies used, therapist gains from training are not sustained without ongoing support, such as consultation or supervision (Herschell et al., 2010).

Techniques Used in EBT Consultation

Research on consultation methods provides guidance on supervision techniques that may impact clinician behavior. Consultants differ from workplace-based supervisors in that they are (1) typically Doctoral-level professionals with expertise in the treatment being studied; (2) supervising clinicians on a subset of their caseload; and (3) focused on the delivery of a particular EBT. A review of consultation identified several consultation techniques that may support high quality EBT delivery, including case review, self-evaluation, feedback, behavioral rehearsals, didactic instruction, and observation (Edmunds, Beidas, & Kendall, 2013a). Studies on EBT consultation have exceeded that of workplace-based EBT supervision and suggest that EBT-focused consultation positively impacts providers' EBT delivery. Studies show that coaching and corrective feedback, based on review of practice, improves clinicians' delivery of EBTs (Miller, Yahne, Moyers, Martinez, & Pirritano, 2004; Webster-Stratton, Reid, & Marsenich, 2014). Consultation that involved modeling and behavioral rehearsal also predicted higher therapist adherence to EBT elements planned in the previous consultation session (i.e., practice concordance) whereas discussion of EBT content in consultation did not predict practice concordance (Bearman et al., 2013).

Gaps in the Literature on Supervision Techniques

These bodies of literature suggest numerous supervision techniques that may be used in workplace-based supervision to support clinicians in delivering EBT. The techniques can be broadly grouped into categories of directive teaching, practice monitoring, experiential

learning, feedback, reflection, and client monitoring. While these provide guidance on techniques that may be used in workplace-based supervision, there are important distinctions that may limit the degree to which these techniques are adopted in workplace-based supervision. The supervisors and consultants in efficacy studies and consultation efforts are often Doctoral-level practitioners with expertise in the particular EBT, as opposed to workplace-based supervisors who are predominantly Master's-level practitioners (Bailin, Bearman, & Sale, 2018; Dorsey et al., 2017). In these studies, supervision is often provided under ideal circumstances, including carefully selected therapists, intensive supervision, and close monitoring (Roth et al., 2010). The literature on manipulations of supervision techniques does not identify the extent to which these supervision techniques occur in workplace-based supervision or their effectiveness in influencing clinician's clinical practice.

Characterizing Techniques in Workplace-Based Supervision

Recent studies have objectively measured the techniques used and content covered during workplace-based supervision (Bailin et al., 2018; Dorsey et al., 2018). One study measured the presence and intensity of supervision techniques used during EBT-focused supervision in 438 supervision sessions with 28 supervisors and 70 clinicians in the context of a state-funded implementation of Trauma-Focused Cognitive Behavioral Therapy (TF-CBT) in 17 community mental health organizations (Dorsey et al., 2018). The most frequently used techniques in supervision, regardless of intensity, were supportive listening (99%), information gathering (97%), didactic instruction (93%), providing clinical suggestions (86%), and fidelity/adherence assessment (64%). Eight of the eleven techniques were used with low intensity. Those that were used with medium to high intensity included supportive listening, information gathering, didactic instruction, and providing clinical suggestions.

Another recent study measured the frequency, duration, and competency of micro-skills (i.e., supervision content and techniques) present in 58 supervision sessions between

13 supervisors and 20 supervisees treating youth with a variety of problem areas in 20 private and public mental health agencies (Bailin et al., 2018). Supervision was focused on clinician treatment delivery broadly, not just EBT delivery. Among supervisory micro-skills considered to be evidence-based (i.e., reference to evidence-based practice elements, corrective feedback, modeling, behavioral rehearsal), their frequency of use across sessions was highly variable, ranging from 70.2% (modeling) to 1.8% (behavioral rehearsal), as was the average session time allocated to these micro-skills (reference to EBT– 20.2% of session time; behavioral rehearsal – 0.3%). The competency with which these micro-skills were delivered was also quite low. Other micro-skills that occurred frequently and were allotted significant time included administrative tasks, case conceptualization, praise, expression of empathy, supervisor self-disclosure, and collaboration.

These descriptive studies identify discrepancies between 'gold standard' techniques and those used in workplace-based supervision. Dorsey and colleagues (2018) found that aside from fidelity assessment and symptom monitoring, common 'gold standard' techniques were used infrequently and/or used mostly with low intensity. These include review of actual practice, behavioral rehearsal, and supervisor modeling. Bailin and colleagues (2018) similarly found infrequent use of role-play but found high frequency use of modeling, though it was used with low competency. Across categories of techniques, the studies found high-to-moderate use of directive techniques (e.g., didactic instruction, recommendations and clinical suggestions), infrequent or poor competence with experiential techniques (e.g., behavioral rehearsal, modeling), and frequent passive supportive techniques (e.g., supportive listening, empathy). The use of monitoring and quality assurance techniques (e.g., symptom monitoring, reviewing actual practice) were mixed, with symptom monitoring being common, but review of practice occurring rarely.

Exploring Supervision-Technique Clusters

These descriptive studies of supervision advance our understanding of the frequency and intensity of specific techniques being used in workplace-based supervision. Theoretical perspectives on supervision suggest that supervisors must combine a variety of experiential learning techniques to influence clinician competency (Milne, 2009). Reviews of studies that manipulate supervision techniques demonstrate that the use of multifaceted supervision approaches often involve teaching, modeling, behavioral rehearsal, and corrective feedback (Falender et al., 2004; James, Milne, & Morse, 2008; Kilminster & Jolly, 2000; Milne et al., 2008; Milne & James, 2000). In addition to understanding which techniques are used in workplace-based supervision, studying the interplay of supervision techniques may provide a more realistic characterization of supervision. Given that supervision does not consist of independent strategies, an approach that captures the varying combinations of techniques employed and the intensity with which they are used may better characterize workplace-based supervision. For instance, a subgroup of supervision that could emerge might involve high intensity directive and experiential learning techniques (e.g., didactic instruction, behavioral rehearsal, and modeling), moderate monitoring and quality assurance techniques (e.g., reviewing practice and symptom monitoring), and low intensity passive supportive techniques (e.g., supporting listening and information gathering). In contrast, another subgroup of supervision styles could involve high intensity use of passive supportive techniques (e.g., supportive listening and information gathering), moderate intensity elicitation of clinician ideas, with low or non-occurrence of monitoring and quality assurance techniques and experiential learning techniques. This approach can provide a holistic characterization of supervision.

Beyond characterizing supervision, an important question is whether particular styles of supervision impact clinicians' EBT delivery. Examining the influence of supervision techniques independently in a regression model would only allow us to understand the

relation between a particular technique and clinicians' EBT delivery, after controlling for the inter-relationships between techniques. However, if these techniques interact with one another to enhance, counteract, or non-linearly relate to clinicians' EBT delivery, these relations would not be captured. Rather than operating independently, supervision techniques are likely used in combination to complement one another. Various combinations of techniques may differ in their influence on clinician's learning and skill development.

Current Study

The current study is based on data from a National Institute of Mental Health funded study of workplace-based clinical supervision of EBT in the context of a State-funded implementation of TF-CBT in community mental health organizations (Dorsey et al., 2013). Data come from the descriptive phase characterizing "supervision as usual" for clinicians delivering TF-CBT prior to the initiation of an RCT comparing two supervision approaches. Behaviorally coded individual supervision sessions were used to examine whether subgroups of 'supervision styles' emerged among supervision techniques. Based on these results, subsequent chapters examined the relation among organizational, supervisor, clinician, and supervision session factors and the identified supervision styles. Finally, we examined whether the identified supervision styles predicted clinicians' fidelity in delivering TF-CBT with the trauma-exposed youth whose treatment was discussed during supervision.

Given the exploratory nature of these analyses, we do not propose hypotheses about the specific clusters of supervision techniques that will emerge. To illustrate broad clusters that could emerge and the utility of the cluster approach in characterizing supervision styles, we provide the following example. We expect supervisors may vary in the intensity with which they use different categories of techniques, including experiential techniques (e.g., behavioral rehearsal; modeling), directive techniques (e.g., didactic instruction; clinical suggestions), monitoring and quality assurance techniques (e.g., symptom monitoring;

reviewing practice; fidelity assessment), elicitation (i.e., eliciting clinician ideas) and passive supportive techniques (e.g., supporting listening; information gathering). For instance, we could see the following patterns of technique use (Figure 1):

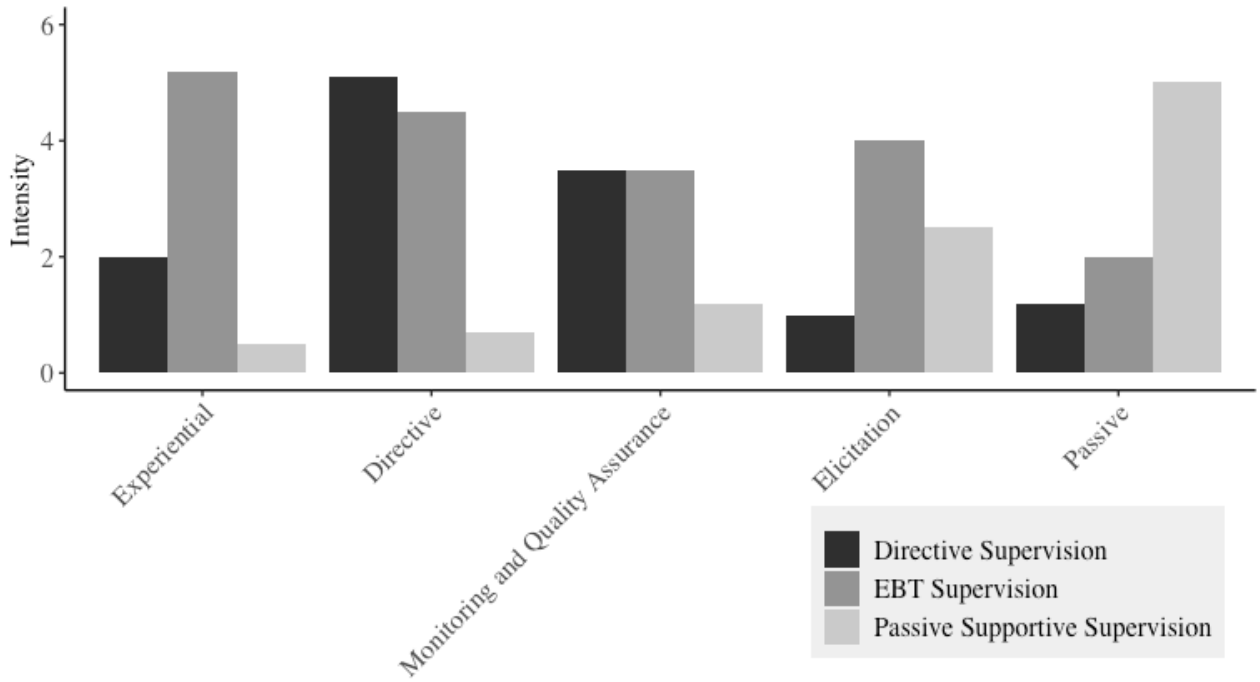
'Directive supervision': low intensity *experiential* techniques, high intensity *directive* techniques, moderate *monitoring and quality assurance*, and low intensity *elicitation* and *passive supportive* techniques.

'EBT supervision': high intensity *experiential* and *directive* techniques, moderate intensity *monitoring and quality assurance* techniques and *elicitation*, and low intensity *passive supportive* techniques.

'Passive Supportive supervision': low or non-occurrence of *experiential* techniques, *directive* techniques, and *monitoring and quality assurance* techniques, moderate intensity *elicitation*, and high intensity use of *passive supportive* techniques.

Figure 1

Sample Supervision Cluster Results



Method

Participants

Table 1 presents demographic information for supervisor and clinician participants.

Supervisors. Participants included 28 supervisors from 17 public mental health organizations located in 23 separate offices. To be eligible to participate, supervisors were required to have TF-CBT-specific training, provided through a Washington State EBT initiative, and to be a current supervisor of two or more study-eligible clinicians. There were no exclusionary criteria. Thirty-three supervisors were enrolled in the descriptive phase, but five did not submit audio recordings of their supervision sessions and were therefore excluded from these analyses. Previous comparisons between those who did and did not submit recordings demonstrated there were few significant differences, except that supervisors who submitted recordings were slightly older (mean age = 44.4 vs. 37.8, $p < .05$), more likely to endorse CBT as their primary theoretical orientation (75 vs. 0%, $p < .05$), and less likely to endorse family systems therapy (21 vs. 60%, $p < .05$) or art/play therapy (0 vs. 40%, $p < .05$) (Dorsey et al., 2018).

Clinicians. Participants included 70 clinicians who were recorded in supervision sessions. Ninety-five clinicians were enrolled in the descriptive phase and among those 70 (74%) were included in the submitted supervision recordings. To be eligible for study inclusion, clinicians must have been trained in TF-CBT through the EBT initiative, provided TF-CBT to children and adolescents, supervised by a participating supervisor, employed at least 80% full-time equivalent, and provided treatment in English (to allow for coding of TF-CBT fidelity). As previously reported (Dorsey et al., 2018), few significant differences emerged: clinicians who were recorded had provided psychotherapy for longer ($M = 7.0$ vs. 4.3, $p < .05$) and were less likely to have a degree in Marriage and Family Therapy (11% vs. 40%, $p < .05$) than those who were not recorded.

Table 1

Supervisor and Clinician Participant Characteristics

Variable	Supervisors (n = 28)		Clinicians (n = 70)	
	n	%	*n*	%
Female	18	64.3%	61	87.1%
Race/Ethnicity				
White/Caucasian	26	92.9%	62	88.6
Hispanic or Latinx	0	0.0%	8	11.4
Asian	1	3.6%	3	4.3
Native Hawaiian/Other	1	3.6%	1	1.4
Black/African American	0	0.0%	–	–
Other	0	0.0%	2	2.9
Education level				
Bachelor's	0	0.0%	5	7.1
Master's	26	92.9%	62	88.6
Doctoral	2	7.1%	3	4.3
Academic degree/background				
Marriage and family therapy	5	17.9%	8	11.4
Psychology	3	10.7%	4	5.7
Social work	11	39.3%	19	27.1
Counseling psychology	9	32.1%	28	40.0
Other	0	0.0%	11	15.7
Primary Theoretical Orientation				
Cognitive-behavioral	21	75.0%	45	64.3
Family systems	6	21.4%	7	10
Solution-focused	1	3.5%	3	4.3
Humanistic	0	0.0%	4	5.7
Psychodynamic	0	0.0%	7	10.0
Play therapy	0	0.0%	3	4.3
Art therapy	0	0.0%	1	1.4
Licensed	27	96.4%	36	51.4
	M	***SD***	***M***	***SD***
Years providing psychotherapy	-	-	7.0	6.2
TF-CBT clients seen in past 3 months	-	-	6.0	7.2
TF-CBT Knowledge	10.43	1.53	9.0	1.9
Years providing supervision	7.71	5.14	-	-
Self-efficacy supervising	3.42	0.54	-	-
EBT implementation climate	-	-	3.1	0.53

Children and adolescents. Participants included 60 children and their guardians who received TF-CBT from an enrolled clinician. To be eligible, children and adolescents were required to be clients at a participating agency. Inclusion criteria included: age 6–17; trauma history; significant posttraumatic stress (PTS) symptoms; live with a parent/legal guardian who is willing to participate in the study; the child is English-speaking; and treatment approach will be TF-CBT. Exclusionary criteria were intentionally limited and included only: presence of a pervasive developmental disorder or cognitive impairment and parental serious mental illness.

Measures

Supervision Techniques. The intensity with which supervisors used 13 supervision techniques during supervision sessions was objectively coded using the Supervision Process Observational Coding System (SPOCS; Dorsey et al., 2018). Table 2 includes a description and example of the supervision techniques. The SPOCS was adapted from the Therapeutic Process Observational Coding System for Child Psychotherapy— Strategies scale (TPOCS-S; Garland et al., 2010; McLeod & Weisz, 2010) a coding measure for measuring psychotherapy strategies in usual care. The SPOCS includes 29 supervision strategies: 16 content areas and 13 techniques. The current study uses data from coded supervision techniques. The supervision techniques were derived from a review of the literature (Hsu & Fan, 2010; Milne et al., 2008; Reiser & Milne, 2017), other supervision and consultation coding manuals (Edmunds, Kendall, et al., 2013; Milne et al., 2011; Nakamura et al., 2014) and expert consensus. Though developed for TF-CBT coding, the SPOCS supervision techniques are likely applicable to supervision of general treatment. Five techniques are considered 'gold standard' techniques based on their inclusion in efficacy trials (Roth et al., 2010), expert consultation (Edmunds, Beidas, & Kendall, 2013b), and training (Beidas & Kendall, 2010): symptom monitoring, reviewing actual practice (audio/video, client work produced in

session), fidelity or adherence assessment, clinician behavioral rehearsal in supervision, and supervisor modeling.

Trained coders rated technique occurrence in 5-min intervals (low, medium, or high), yielding intensity scores for techniques for the session (0–6 range; 0: *nonoccurrence*; 1–2: *low*; 3–4: *medium*; 5–6: *high intensity*). For instance, a low-intensity rating of supportive listening would be given for a limited number of supervisor non-specific acknowledgements or general praise (e.g., "nice work"; "that sounds hard"), while a higher score would be given if the supervisor provided more frequent and explicit support, validation, or praise (e.g., "…sounds like a tough session; still, you did a really nice job getting this super anxious kid to feel comfortable talking about his sexual abuse. I am impressed."). Audio recordings were used for coding, resulting in coding of only verbal behavior. Supervision sessions were submitted by supervisors weekly throughout a one-year study period.

Twenty-four percent of the coded supervision session recordings were coded by multiple coders to assess interrater reliability. The overall group average ICC was ICC(2,6) = 0.87, suggesting excellent reliability (Cicchetti, 1994). Individual coders had excellent ICCs of 0.84 or higher. At the item level, ICCs ranged from 0.19 to 0.95. Of note, only 2 of the individual 13 item-level codes were below 0.60. The one in the "poor" range (< 0.40), Assigning Additional Training/Learning, had relatively low incidence and low variance, which can result in unreliable estimates of interrater reliability (Banerjee, Capozzoli, Mcsweeney, & Sinha, 1999; Cicchetti & Feinstien, 1990; Hutchinson, 1993). Reviewing Assigned Suggestions or Trainings had an ICC of 0.47, reflecting "fair" reliability, and also had low incidence and variance. Given the poor reliability and low incidence, these two techniques were excluded from the analyses.

Table 2

Supervision Process Observational Coding System (SPOCS): Supervision Technique Domains

Techniques	
Symptom Monitoring	Supervisor and/or clinician discuss repeated use of standardized assessment measures to determine how the child and/or caregiver(s)' symptoms look currently and how they have changed over time, e.g., to examine treatment effectiveness. *E.g.* "So tell me again, he's at a 13 now, what was his score on the CPSS when you started?"
Progress Note Review	Supervisor reviews the progress note with clinician or alludes to review that occurred before the supervision meeting. *E.g.* "Let's look at your case note for that session".
Providing Clinical Suggestions	Supervisor gives specific ideas, suggestions, and/or directions to clinician about what to do in a future session and/or about what clinician should have done in a past session. *E.g.* "You said that last session the mom asked about what happens at court. I might have walked through the process or given her ideas for how to find out more information."
Assigning Additional Training/Learning	Supervisor makes a clear request for clinician to obtain additional training or expertise for his/her own learning (does not have to directly benefit the case being discussed). *E.g.* "There's a chapter in the new TF-CBT book on doing TF-CBT with kids in foster care. You should check that out and see if there are any ideas for this client."
Reviewing Assigned Suggestions/Training	Supervisor specifically checks in about and/or asks about a suggestion, strategy, training, and/or other recommendation from a past supervision session. *E.g.* "So, help me understand, it sounds like you didn't sign up for the CBT+ training like we talked about. What happened with that?"
Review of Actual Practice	Prior to or during supervision, the supervisor either: a) watched or listened to a recording of clinician's actual clinical practice with a client or b) reviewed client work from a past TF-CBT session, e.g., reviews the child's trauma narrative, looks at triangles the child and/or caregiver(s) created during cognitive coping. *E.g.* "Watching part of your tape, I noticed that you did most of the talking throughout the session. Tell me why that might not be the most effective way to change the mother's behavior with her child."

Fidelity/Adherence Assessment	Supervisor and/or clinician discuss the topic of fidelity to TF-CBT or clinician's progress through the model. *E.g.* "As you know, PRACTICE is an acronym for the TF-CBT components. Tell me which components you have completed and which one you're on now."
Elicitation	Supervisor uses questions to a) encourage/elicit clinician thinking and planning for a subsequent session (as opposed to providing ideas/suggestions) or b) help clinician evaluate his/her own effectiveness in a past session. *E.g.* "He's really blaming himself for what happened. What do you think are some possible other ways to view the situation, given what you know about the case?"
Didactic Instruction	Supervisor provides information, teaches, and/or explains something to clinician via "lecture" or in a didactic style. *E.g.* "There is research showing that sometimes PTSD symptoms can sound like hallucinations, for example thinking they can hear the offender talking."
Supportive Listening	Supervisor makes statements that reflect, validate, acknowledge, and/or praise clinician. E.g. Supervisor: "Sounds like a frustrating situation."
Clinician Behavioral Rehearsal (in supervision)	Supervisor guides clinician through practicing effective use of a therapeutic skill or technique for a future session; clinician actively practices a skill playing the role of the therapist. E.g. "Okay, let me just pretend to be the kid—'but it's too hard! I can't talk about what happened.'" [Clinician proceeds to play the clinician role and respond]
Supervisor Modeling	Supervisor models (i.e., enacts or demonstrates) a specific clinical skill or method of delivering a treatment component. *E.g.* "You might say, 'Hey, is there anything that happened in your past that guides how you deal with your daughter?'"
Information Gathering	Supervisor gathers information about the case, a past session, and/or clinician's therapeutic/TF-CBT skill-level, e.g., asking the clinician for information about the child, child's symptoms, family's problems, and/or context, e.g., living environment. E.g. "When is the IEP meeting?"

Procedures

Audio recordings of supervision were collected during the descriptive phase of the two-phased parent study, which characterized "supervision as usual" for clinicians delivering TF-CBT prior to the initiation of an RCT comparing two supervision approaches. The parent study builds on a statewide EBT training initiative. In 2007, Washington State began modestly funding training in TF-CBT for public mental health organizations. Beginning in 2009, training also included CBT for depression, anxiety, and behavior problems, with 100–250 trainees per year. The training program included a 3-day in-person learning session that was active (i.e., opportunities to practice and obtain feedback) and six months of twice-monthly 1-hour expert telephone consultation on applying the CBT model to trainees' caseloads. Supervisors and clinicians had at least 1 or 2 training cases, respectively. Optional monthly technical assistance calls and a yearly one-day supervisor training were made available to supervisors. Each organization was required to have at least one supervisor complete the initiative expectations. Organizations were able to send trainees annually to address organizational growth and attrition. As of 2015, 83% of the 109 public mental health organizations in Washington State had participated in at least one training.

Procedures were approved by the Washington State Institutional Review Board. Supervisors and clinicians provided informed consent to participate. Supervisors audio recorded the portions of their supervision sessions that pertained to enrolled clinicians and their TF-CBT cases over the course of one year (October 2012 – September 2013). Audio recordings were submitted to the research team for coding. Potential participants were first identified through organizations that had participated in the state-funded EBT initiative and still had at least one TF-CBT-trained supervisor in their organization. Upon agreeing to participate, supervisors identified potentially eligible clinicians among those they currently supervised. Clinicians were invited to participate in the study by the research team and

informed consent was obtained. If interested, study staff proceeded with recruitment. Enrolled clinicians were asked to introduce the study to caregivers of all youth who were potentially eligible to receive TF-CBT and who met the study inclusion and exclusion criteria. If families were interested, study staff proceeded with informed consent and recruitment.

Intervention

Supervisors in the current study were supervising clinicians in their delivery of TF-CBT with trauma-exposed youth. TF-CBT is an EBT that applies principles of cognitive-behavioral therapy to address symptoms of posttraumatic stress, depression and disruptive behaviors among children exposed to traumatic events (Cohen, Mannarino, & Deblinger, 2006; Dorsey, McLaughlin, et al., 2017). It consists of the following components: psychoeducation, parenting (for caregivers), relaxation, affective modulation, cognitive coping, trauma narrative, cognitive processing, in-vivo exposure, conjoint parent-child sessions, and enhancing safety. TF-CBT is a conjoint parent-child treatment, designed to be delivered in 12-16 sessions. Parents and children engage in parallel sessions for each treatment component and conjoint sessions, some of which involve the child sharing their trauma narrative with their caregiver in the session. Caregivers may also join sessions with the child focused on enhancing the child's safety.

Data collection

For one year (October 2012–September 2013), supervisors were asked to audio record the portions of their individual supervision sessions that included discussion of participating clinicians' TF-CBT cases and submit these to the research team. Informal and group supervision sessions were not recorded. The supervision audio recordings were saved on study-provided, password-protected tablet devices. Recordings were transferred to the research team using a cloud-based server compliant with the Health Insurance Portability

and Accountability Act of 1996. Organizations were compensated $3,000 for their participation at the end of the second phase of the RCT.

Coder training

Coders were six post-baccalaureate research assistants. They were first trained to reliably code TF-CBT treatment sessions (described in the Chapter 3 method section), given that study investigators assumed ability to code TF-CBT fidelity was a prerequisite for coding TF-CBT supervision. Coders attended a 2-day TF-CBT clinical training, completed a 10-hour TF-CBT online course, read the TF-CBT treatment manual (Cohen et al., 2006), and participated in didactic training in distinguishing components of the treatment model with a TF-CBT treatment developer, expert trainer (and study PI), and an experienced coder in prior TF-CBT studies. Coders were also trained to code supervision of TF-CBT. Coder training for fidelity and supervision coding consisted of similar activities, involving independent review of the respective manuals (TPOCS and SPOCS manuals), didactic training, independent coding, group review of coded sessions, and joint listening to sessions when necessary to reach consensus. Three expert coders first coded 10 training files and came to consensus on their codes. Coders then independently coded 10 training files to ensure acceptable inter-rater reliability across group members and with the expert trainers. Coder training was complete once their individual ratings at the overall level met the threshold for inter-rater reliability, intra-class correlation coefficient (ICC) (2,1) ≥ .80 (Cicchetti, 1994). If an individual content or technique item had an ICC (2,1) ≤ .60, coders were assigned additional review and practice. Coders were required to review the respective coding manuals monthly and to attend periodic booster trainings to maintain high reliability. Coders were also periodically provided with feedback on their inter-item reliability. Audio recording files were randomly assigned to each coder.

Session sampling

A total of 667 supervision recordings were received from 28 supervisors who submitted individual TF-CBT supervision sessions. All files were included, with the exception of 29 files shorter than 1 min (4.3%). Of the remaining 638, 438 (70%) were coded. Twenty-three recordings were coded per supervisor. When supervisors submitted greater than 23 recordings, stratified random sampling was used to select 23 recordings that were spread across time and clinicians. Eighteen supervisors submitted fewer than 23 recordings, and all were coded (M= 10.8; SD =4.9; range 4–19).

Analytic plan

A hierarchical cluster analysis utilizing minimum variance method (i.e., Wards method) with a Canberra distance measure was conducted to identify patterns of supervision techniques across eleven objectively coded supervision techniques used during clinical supervision. The Canberra distance measure was selected due to its appropriateness for ranked outcome measures (Jurman, Riccadonna, Visintainer, Furlanello, & Fondazione, 2009). The purpose of cluster analysis is to identify homogenous subgroups based on shared properties. Cluster analysis is a data reduction technique, such as factor analysis and discriminant analysis, which aims to increase within-group homogeneity and between-group heterogeneity (Landau & Chis Ster, 2010; Norusis, 2010). It differs from other data reduction techniques in that it groups together observations that are close together in a multidimensional space to identify homogenous groups without previously specifying the group membership or the number of groups. In the current study, an observation represents supervision techniques for a given supervisor-clinician dyad during a given session. The agglomerative hierarchical method was used because it does not require the specification of the number of clusters, which is more appropriate for exploratory analyses. In accordance with recommended methods, data were standardized before conducting cluster analysis

(Everitt, Landau, & Leese, 2001). A dendrogram, displaying the cluster hierarchy as a tree diagram, and a scree plot of the agglomerative coefficients were inspected to aid in selection of the demarcation point to identify an appropriate cluster solution.

Cluster Membership

Cluster assignments represent supervision-technique clusters for a given supervisor, with a given clinician, during one supervision session. Clustering of observations at this level allowed for the measurement of variability in supervision-technique clusters within a supervisor-clinician dyad across their supervision sessions. Clustering was then summarized at two levels: 1) the proportion of sessions a supervisor was assigned to a particular cluster with a *given* clinician (i.e., supervisor-clinician level summary) and 2) the proportion of sessions for which a supervisor was assigned to a particular cluster across *all* of their clinicians (i.e., supervisor-level summary). This approach allowed for the characterization of the variability in cluster membership at both the supervisor-clinician dyad and supervisor level.

Fit Measures

The agglomerative coefficient and cophenetic correlation for the overall cluster analysis were evaluated. The agglomerative coefficient measures the dissimilarity of an object to the first cluster it joins, divided by the dissimilarity of the final merger in the cluster analysis, averaged across all samples. Values closer to 1 suggest a more balanced clustering structure. The cophenetic correlation is a multiple correlation between input dissimilarities and the output dissimilarities implied by the resulting dendrogram. It measures how well the dendrogram represents the dissimilarities among observations. Values closer to 1 are better.

Validation Measures

Internal validation measures were used to assess the goodness of the clustering structure. Validation measures assessed the compactness and separation of the clusters

(Everitt et al., 2001). *Compactness* measures how close observations are within the same cluster (i.e., cluster homogeneity). *Separation* measures the degree of separation between clusters. The Silhouette Width estimates the average distance between clusters with values closer to 1 indicating well clustered observations and negative values indicating observations are likely placed in the wrong cluster. The Dunn Index assesses both compactness and separation, with larger values indicating more compact, well-separated clusters.

To assess the stability of the cluster solution, a nonparametric bootstrapping approach (Hennig, 2007) was used to draw random samples with replacement (i.e., some data points may be sampled more than once) from our data (N = 438, *B* samples = 1000). The cluster analysis was conducted on each of the generated samples and the similarity between the original and each new clustering was calculated. To assess the stability of an original cluster to a cluster in the bootstrapped sample, a similarity measure was used to assess set membership (i.e., cluster membership).

Cluster similarity was assessed using the Jaccard coefficient, a measure of the similarity of two sets by calculating the proportion of shared observations that are assigned to the same cluster divided by the total number of shared observations. For every original cluster, the most similar cluster in each bootstrapped sample was identified using the maximum Jaccard Coefficient. If the maximum Jaccard coefficient (ranging from 0 – 1) is less than 0.5, a cluster is considered dissolved (i.e., it did not reappear in the simulated sample). The mean maximum Jaccard coefficient for each cluster was calculated across the 1000 samples to estimate the stability of each cluster. Clusters with stability values of 0.85 are considered highly stable, values between 0.6 and 0.75 are considered to be meaningful but lacking high certainty of which points should be clustered together, and values below 0.6 are considered unstable. For clusters that were unstable, the bootstrapping approach was conducted on the cluster solution with the nearest agglomeration coefficient to evaluate

whether an alternative cluster solution was more stable. Compactness and separation were also recomputed on the new cluster solution and compared to the original results.

Results

Results for the overall cluster analysis showed a high agglomerative coefficient (0.97), suggesting a balanced clustering structure, but low cophenetic correlation (0.48), suggesting a poor fit between the original data and the clustering. Upon inspection of the dendrogram and scree plot (Figure 2), a two-cluster solution appeared to fit the data best and the internal validity and stability measures for a two-cluster solution were examined (Table 3). The internal validity and stability metrics for the two-cluster solution were poor, therefore, the dendrogram and scree plot were re-inspected, and the internal validity and stability metrics were calculated for cluster solutions ranging from three to seven as these cluster solutions incrementally reduced the within-group variance, reflecting more compact clusters. While the seven-cluster solution produced better within-cluster compactness and between-cluster separation, the cluster stability based on the bootstrapped Jaccard Coefficient was poor.

Figure 2

Dendrogram and Scree Plot of Hierarchical Cluster Analysis

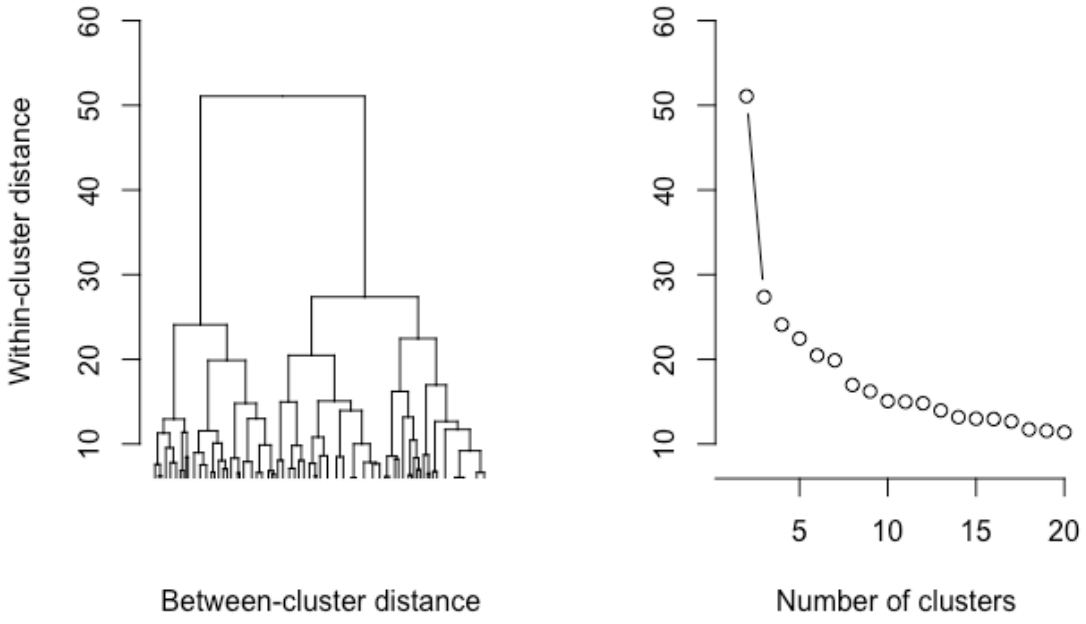


Table 3

Model 1 Hierarchical Cluster Analysis of All Supervision Techniques

Cluster N[a]	Silhouette Width[b]	Dunn Index[c]	Jaccard Coefficient[d]
2	0.20	0.09	0.58 – 0.66
3	0.11	0.09	0.42 – 0.64
4	0.09	0.09	0.47 – 0.67
5	0.08	0.10	0.37 – 0.62
6	0.08	0.10	0.30 – 0.58
7	0.21	0.13	0.31 – 0.55

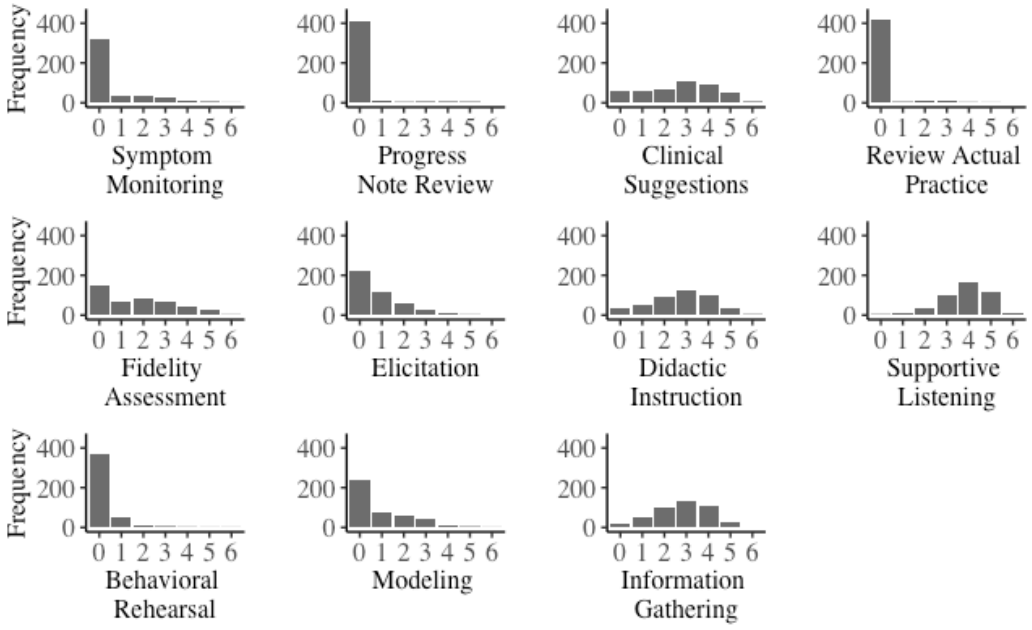
[a]Cluster N refers to the number of clusters in each cluster-solution.
[b]Silhouette Width calculates the average distance between clusters. Values closer to 1 indicate well clustered observations, negative values indicate inaccurate clustering.
[c]Dunn Index assesses cluster compactness and separation. Larger values indicate more compact, well-separated clusters.
[D]Jaccard Coefficient estimates cluster stability. Values above 0.75 are considered stable and value of 0.85 or higher are considered highly stable.

Rationale for data reduction

Exploratory methods, such as exploratory factor analysis, aim to reduce data dimensions to those that are most meaningful and are guided by theory and the observed data (Hogarty, Kromrey, Ferron, & Hines, 2004). We took a similar approach to iteratively reduce the supervision techniques to improve the cluster solution goodness-of-fit. The inclusion of unnecessary or non-informative variables can degrade the results of cluster analysis (Raftery & Dean, 2006). Techniques with extremely low occurrence and variability may lack meaningful information to differentiate between supervision styles. Descriptive statistics for the supervision techniques revealed several techniques occurred very rarely and had restricted variability (Figure 3). Therefore, we adopted an approach that took into account both the theoretical importance of techniques and the variability observed in the sample, removing techniques with the most limited variability and those with less theoretical relevance. Given our aim to understand how supervisors support EBT delivery, our goal was to retain techniques we theoretically believed to impact EBT delivery and also that had reasonable variability.

Figure 3

Supervision Technique Score Frequency Counts



Three techniques, use of a fidelity assessment, elicitation, and supervisor modeling, were also considered theoretically important and had only moderately limited variability. Supervisor modeling has been shown to predict greater clinician adherence to the treatment plan (Bearman et al., 2013) and supervision that included a focus on adherence to treatment principles predicted greater clinician treatment adherence (Schoenwald, Sheidow, & Chapman, 2009). Although supervisor elicitation of clinician ideas and self-reflection has not been studied as often, formative models of supervision based on learning theory suggest that supervisor elicitation of self-reflection is important for developing clinical competency (Milne, 2009). Moreover, CBT therapists report that supervisor elicitation of self-reflective practice is among one of the most helpful techniques in developing procedural skills in therapy (Bennett-Levy, McManus, Westling, & Fennell, 2009). These three techniques were therefore retained in the subsequent cluster analysis.

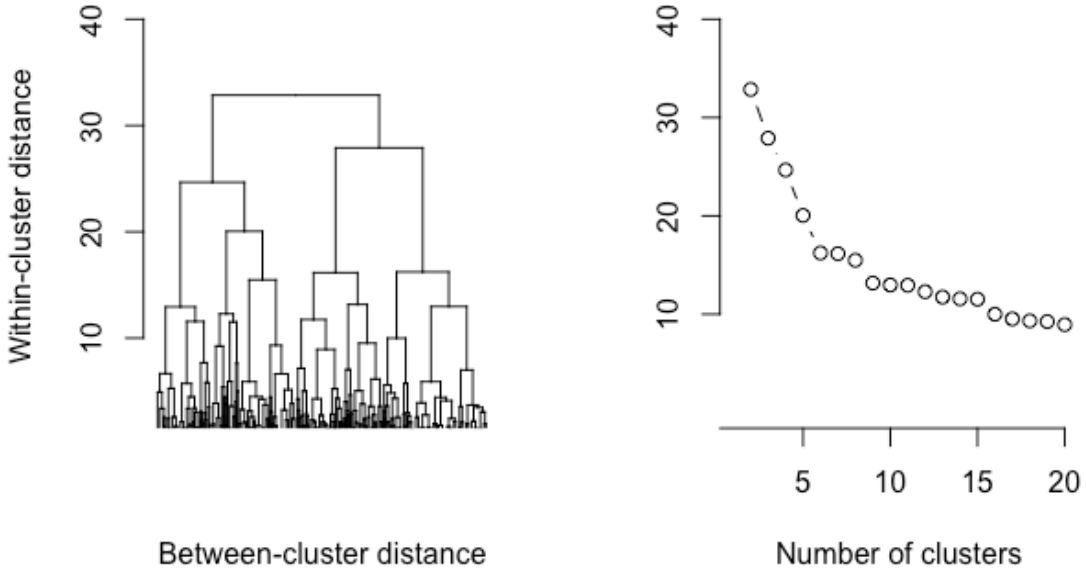
Given the exploratory nature of the analyses and incorporation of a data-driven and theoretical method for variable reduction, the results of the cluster analysis should be interpreted with caution. Another method for evaluating the validity of cluster results is assessing whether clusters are associated with relevant constructs in a hypothesized direction (Shalizi, 2009). Examining the relation among cluster solutions and variables in chapters 2 and 3 provides another opportunity to explore the external validity of the cluster solutions.

Hierarchical cluster analysis with eight techniques

A second cluster analysis was conducted including eight of the eleven supervision techniques: symptom monitoring, clinical suggestions, fidelity assessment, elicitation, teaching, supportive listening, modeling, and information gathering. Results showed a high agglomerative coefficient (0.97) but low cophenetic correlation (0.49), again indicating a balanced clustering structure but a poor fit between the original data and the clustering.

Figure 4

Dendrogram and Scree Plot of the Eight-Item Hierarchical Cluster Analysis



Inspection of the dendrogram and scree plot suggested a four-cluster solution may fit the data best (Figure 4). However, the internal validity and stability indices showed a poor fit (Table 4). In particular, the Jaccard coefficient indicated the clusters were unstable. Three and two-cluster solutions were evaluated next. The three-cluster solution also showed unstable clustering, but the two-cluster solution demonstrated good stability of the clustering results as evidenced by the Jaccard Coefficient. However, given the low cophenetic correlation, indicating a low goodness-of-fit for the overall clustering, this model was not retained.

Table 4

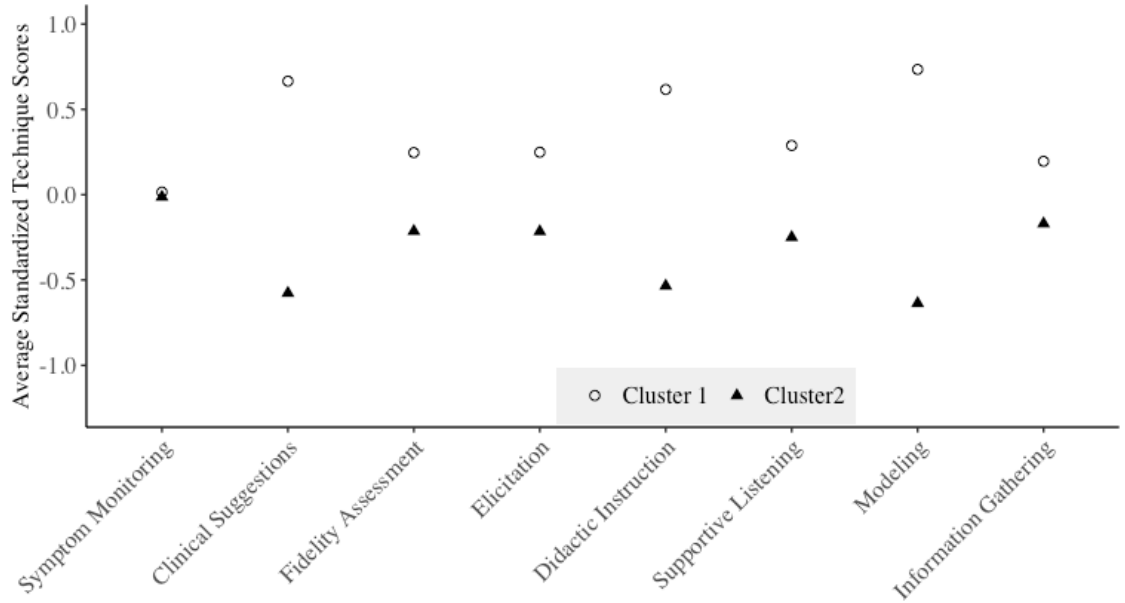
Model 2 Hierarchical Cluster Analysis of All Supervision Techniques

Cluster N[a]	Silhouette Width[b]	Dunn Index[c]	Jaccard Coefficient[d]
2	.21	.13	0.79 – 0.83
3	.14	.12	0.44 – 0.77
4	.12	.12	0.47 – 0.71

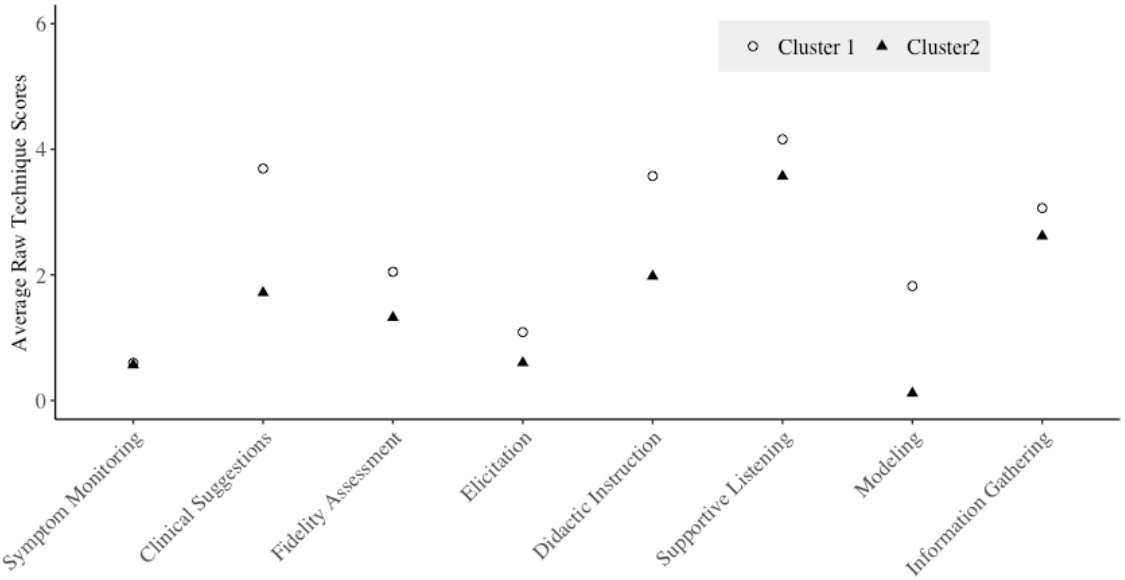
Figure 5

(a) Average Standardized Supervision Technique Scores for the Eight Item Two-Cluster Model
(b) Average Raw Supervision Technique Scores for the Eight Item Two-Cluster Model

(a)



(b)



The two-cluster solution was visually inspected. Figure 5 shows the average score on each technique for observations assigned to cluster 1 and 2. Observations assigned to cluster 1 showed higher scores across all techniques; however, the degree of separation varied by technique. Symptom monitoring shows very low separation in scores for observations assigned to cluster 1 and 2. Similarly, supportive listening and information gathering showed less than a 0.50 *SD* separation between observations assigned to clusters 1 and 2.

Rationale for reducing from eight techniques to five techniques

When symptom monitoring is completed routinely and used to augment treatment it can improve outcomes over usual care (Lambert et al., 2006). However, clinicians have reported challenges to routinely monitoring client outcomes (Garland, Kruse, & Aarons, 2003) and supervisors' discussion of *any* client data only occurred in 12% of coded supervisions sessions (Bailin et al., 2018). Despite symptom monitoring being a potentially impactful supervision technique, the low rates of use and poor separation in the cluster analysis suggest that supervisors in community mental health may not meaningfully differ in their use of symptom monitoring. Additionally, while supportive listening and information gathering showed good variability across supervision sessions, they did not demonstrate good discrimination between the clusters, potentially reducing the internal validity indices. Given that these two techniques reflect very general clinical and supervision skills, they may not be specific enough to differentiate between styles of supervision. Therefore, we recalculated the cluster analysis excluding symptom monitoring, supportive listening, and information gathering.

Hierarchical cluster analysis with five techniques

A third cluster analysis was conducted including five of the eleven supervision techniques: clinical suggestions, fidelity assessment, elicitation, teaching, and modeling. Results for the overall cluster analysis for the third model show a high agglomerative

coefficient (0.99) and an acceptable cophenetic correlation (0.64; Everitt et al., 2001), suggesting a balanced clustering structure and reasonable fit between the original data and the clustering. Examination of the dendrogram and scree plot (Figure 6) suggests a two-cluster solution may fit the data best. Table 5 shows acceptable internal validation metrics and good stability of the two-cluster solution.

Figure 6

Dendrogram and Scree Plot of Hierarchical Cluster Analysis with Five Techniques

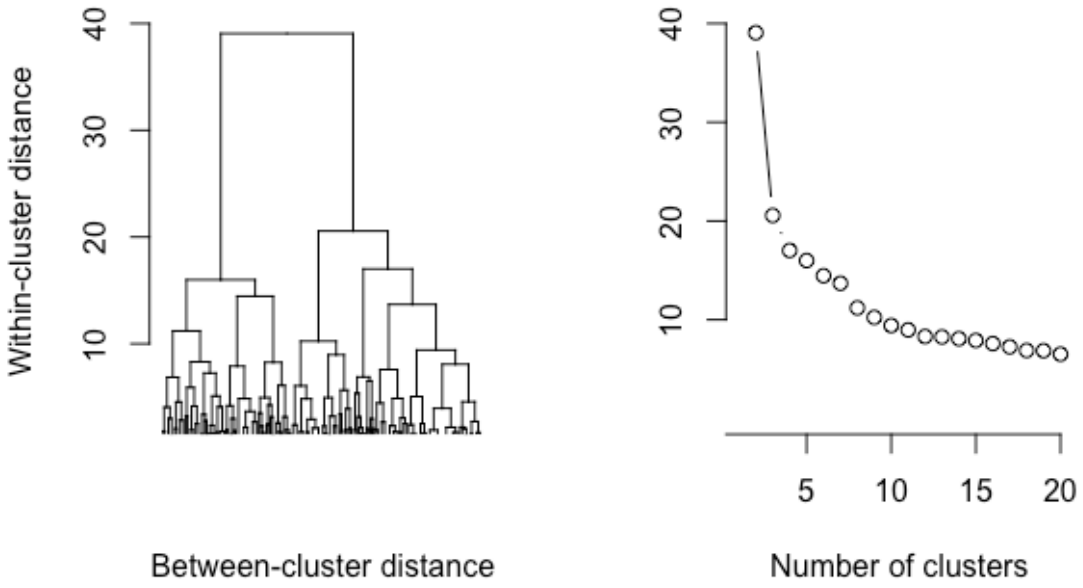


Table 5

Hierarchical Cluster Analysis with Five Supervision Techniques[a]

Cluster N	Silhouette Width	Dunn Index	Jaccard Coefficient
2	0.30	0.17	0.78 – 0.82

[a]Supervision techniques included clinical suggestions, fidelity assessment, elicitation, teaching, modeling

The scatterplots in Figure 7 show that cluster 1 is characterized by higher intensity scores across all techniques, with the greatest differences seen in supervisors providing clinical suggestions and modeling of a particular skill or method. We term this cluster "*directive*" supervision and term cluster 2 "*undifferentiated low*" supervision. Table 6 shows mean scores for the supervision techniques that did not reliably distinguish between the clusters. While these were not included in the final clusters, supervision sessions may have included some use of these techniques, therefore we provide the average scores by cluster and across the sample to fully characterize supervision sessions.

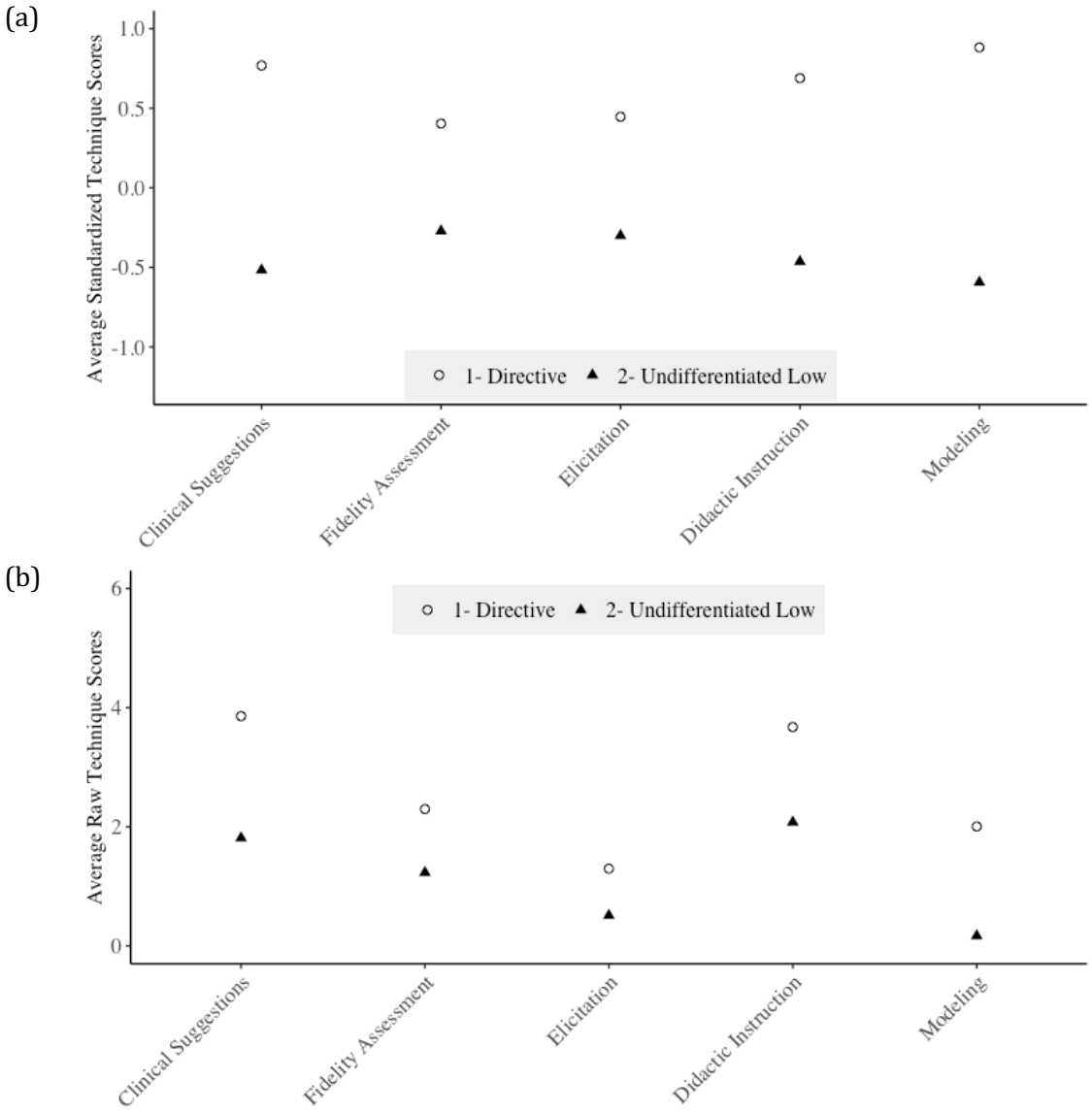
Table 6

Average Technique Scores for Techniques Excluded from the Cluster Analysis

	"Directive" **M (*SD*)**	***"Undifferentiated Low"*** **M (*SD*)**	**Full Sample M (*SD*)**
Behavioral Rehearsal	0.44 (0.93)	0.12 (0.45)	0.25 (0.70)
Review of Actual Practice	0.09 (0.46)	0.14 (0.65)	0.12 (0.58)
Progress Note Review	0.18 (0.79)	0.15 (0.70)	0.16 (0.73)
Symptom Monitoring	0.70 (1.3)	0.50 (1.04)	0.58 (1.15)
Information Gathering	3.15 (1.13)	2.61 (1.23)	2.83 (1.22)
Supportive Listening	4.21 (0.99)	3.60 (1.09)	3.84 (1.09)

Figure 7

(a) Average Standardized Supervision Technique Scores for the Five Item Two-Cluster Model. (b) Average Raw Supervision Technique Scores for the Five Item Two-Cluster Model

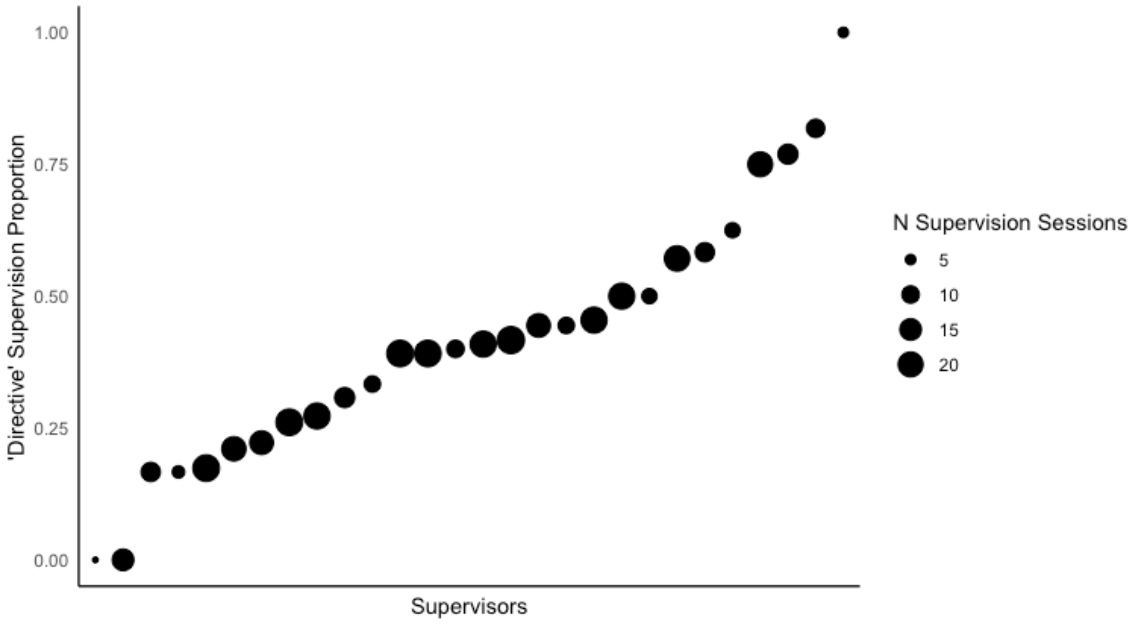


Across all supervision sessions, 40.23% of sessions were assigned to the "*directive*" cluster and 59.77% were assigned to the "*undifferentiated low*" cluster, suggesting that supervision sessions were more commonly characterized by less intensive use of the five supervision techniques. Figure 8 shows the proportion of sessions each supervisor was grouped into the "*directive*" cluster across all of their clinicians. Figure 9 shows the proportion of sessions each supervisor-clinician dyad was grouped into the "*directive*" cluster

across all supervision sessions. The diameter of the dots on the figures shows how many supervision sessions were submitted by each supervisor or supervisor-clinician dyad. At the supervisor level, the average proportion of sessions a supervisor engaged in "*directive*" supervision was 0.41 (*SD* = 0.24). Figure 8 shows that rather than engaging in a particular supervision style across all sessions (i.e., a high prevalence of 0 or 1), supervisors tended to show a mix of "*directive*" and "*undifferentiated low*" supervision across supervision sessions (i.e., a high prevalence of observations around 0.50). The diameter of the observations reflects the number of supervision session files submitted by a supervisor. It does not appear that the number of supervision sessions submitted greatly biased the proportion of sessions supervisors were clustered into "*directive*" supervision.

Figure 8

Proportion of Supervision Sessions Supervisors were Clustered in "Directive" Supervision

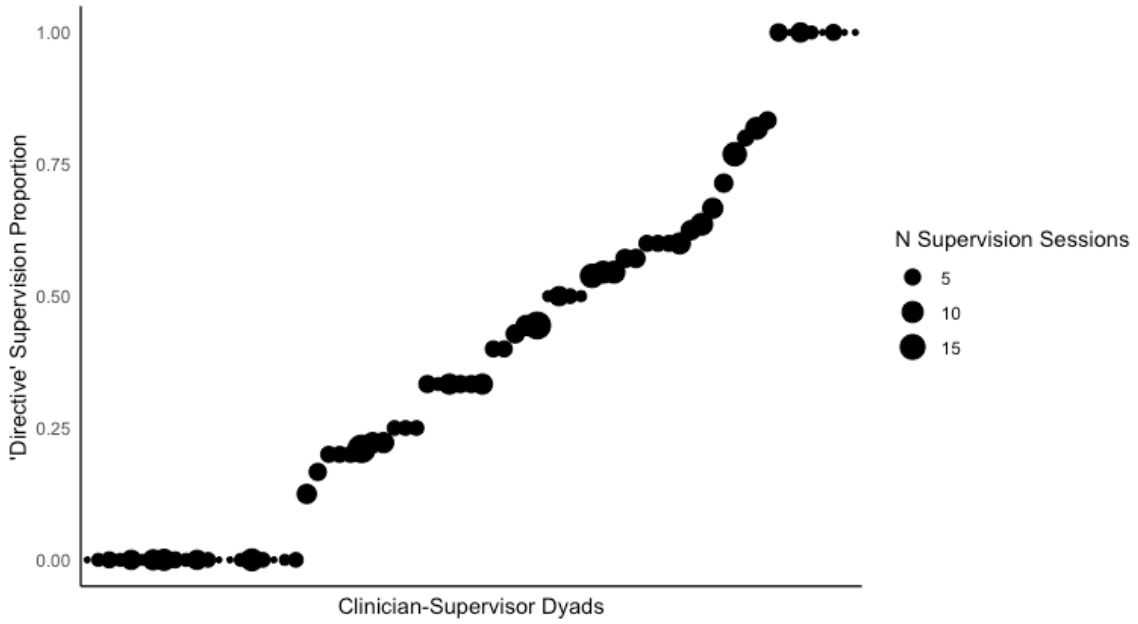


At the supervisor-clinician dyad level, the average proportion of sessions a supervisor engaged in "*directive*" supervision was 0.39 (*SD* = 0.33). Figure 9 shows that there was a greater tendency for supervisors to be clustered into either "*directive*" or "*undifferentiated*

low" supervision when supervising a particular clinician. That is, there is a greater frequency of proportions of 0 and 1. This also seems to be unbiased by the number of supervision sessions represented in the data for a particular supervision-clinician dyad. These figures suggest that supervisors are likely varying their supervision to match the clinician they are working with and tend to be more consistent in their supervision style with a particular clinician than across clinicians.

Figure 9

Proportion of Supervision Sessions Supervisor-Clinician Dyads were Clustered in "Directive" Supervision



Interim Summary

The results revealed two supervision session clusters. The cluster termed "*directive*" was characterized by moderately intensive clinical suggestions and didactic instruction, low-to-moderately intensive fidelity assessment, and low intensity modeling and elicitation. The other cluster, termed "*undifferentiated low*" was characterized by low intensity didactic instruction, clinical suggestions, and fidelity assessment, and generally no elicitation or modeling. No stable clusters were identified when including six other supervision techniques

in the analyses: symptom monitoring, progress note review, review of actual practice, supportive listening, behavioral rehearsal, and information gathering.

Rather than engaging in a consistent style or cluster for all supervision sessions across clinicians, supervisors tended to show a mix of both clusters. Within a particular supervision-clinician dyad, there was a greater tendency for supervisors to consistently engage in one of the two supervision styles, with more dyads demonstrating "*directive*" or "*undifferentiated low*" supervision across all supervision sessions. This suggests that supervisors tended to tailor their supervision depending on the clinician they were supervising.

Among the six excluded techniques, four were believed to be theoretically important in supporting EBT delivery but occurred rarely and did not produce reliable clustering results. These four techniques included symptom monitoring, progress note review, review of actual practice and behavioral rehearsal. Their average intensity scores all fell below a "low intensity" score, both when comparing the mean score across the entire sample and by cluster. The two remaining techniques, information gathering and supportive listening, had less theoretical relevance to EBT delivery specifically, but are likely important techniques in supervision of any treatment. They were used on average with moderate intensity but did not distinguish between the clusters.

Despite the inability to distinguish between supervision clusters when using the full set of eleven techniques, the resulting clusters in the third model produced stable, interesting, and potentially meaningful distinctions between the supervision sessions. One method of validating the results of cluster analysis is to assess whether clusters are associated with relevant constructs in a hypothesized direction (Shalizi, 2009). Therefore, the following chapters examine predictors or supervision clusters and the relation among supervision clusters and TF-CBT delivery. These findings can provide further evidence of the validity of the identified clusters.

CHAPTER 2
PREDICTORS OF SUPERVISION STYLES

Introduction

The findings in Chapter 1 revealed two distinct supervision clusters used during supervision of EBT delivery in community mental health settings. "*Directive*" supervision was characterized by predominantly directive methods, some use of experiential learning and fidelity assessment, and low use of elicitation. "*Low undifferentiated*" supervision consisted of low use of directive methods and fidelity assessment, and nearly no use of elicitation or experiential learning methods. In order to better understand EBT supervision in community mental health settings, we aimed to explore factors that may influence the use of each of these forms of supervision.

Possible predictors of supervision

While there is a growing body of research that has examined the effect of supervision practice on clinical practice (e.g., Alfonsson, Parling, Spännargård, Andersson, & Lundgren, 2018; Martino et al., 2016; Milne et al., 2008; Schoenwald et al., 2009), less is known about factors that influence supervision practice. There are a few recent studies that inform our hypotheses about characteristics of the community mental health organizations, supervisors, clinicians and supervision sessions that may influence supervision practices. Though it should be noted that most of this research has relied fully or in part on the sample of supervisors and clinicians included in the current study (i.e., Dorsey, Pullmann, et al., 2017; Lucid et al., 2018; Pullmann et al., 2018).

Organizational characteristics

EBT Implementation climate is the shared perception of the extent to which EBTs are expected, supported, and rewarded in an organization (Weiner, Belden, Bergmire, & Johnston, 2011). If an organization expects, supports and rewards EBTs, we expect that supervisors will be more likely to use techniques that have been shown to support EBT delivery in research, such as the higher intensive techniques shown in "*directive*" supervision.

EBT implementation climate has positively predicted clinician-reported and objectively coded supervision practices. Specifically, having a more positive EBT implementation climate predicted greater clinician report of time allotted to EBT activities in supervision (Dorsey et al., 2017), greater clinician-reported EBT supervision intensity (Lucid et al., 2018) and greater intensity coverage of EBT content elements (Pullmann et al., 2018). We hypothesized that higher EBT implementation climate would predict greater odds of engaging in "*directive*" supervision, the supervision cluster with higher intensity use of modeling and fidelity assessment, techniques that have been shown to support EBT delivery (Bearman et al., 2013; Schoenwald et al., 2009).

Supervisor characteristics

Variability in the content supervisors cover and techniques they use in supervision can be attributed to both clinicians and supervisors (Dorsey et al., 2018). Among supervisors, it is conceivable that their theoretical orientation, general experience supervising, EBT knowledge and beliefs about their self-efficacy in supervising a particular EBT may impact their supervision practices.

Just as clinician's theoretical orientation has influenced the strategies they use in treatment (Brookman-Frazee, Haine, Baker-Ericzén, Zoffness, & Garland, 2010), supervisors' theoretical orientation may predict their use of supervision techniques. In a bivariate model, Pullmann and colleagues (2018) found that supervisors with a CBT orientation covered assessment with greater intensity than those with a family systems orientation, though this was non-significant after accounting for the influence of EBT implementation climate. Another study did not find an association between supervisor orientation and the use, duration, or competency of objectively coded EBT content coverage or evidence-based supervision techniques (i.e., corrective feedback, modeling, role-play; Bailin et al., 2018). Despite this mixed evidence, supervisors who endorse a cognitive-behavioral orientation

may be more comfortable with and skilled in the techniques that differentiate between the clusters due to the inclusion of many of these techniques in CBT interventions. For instance, CBT interventions include didactic instruction when providing psychoeducation and when teaching a new skill (e.g., Beck, 2011; Cohen et al., 2006). Therapists tend to engage in modeling when introducing a new skill, such a modeling relaxation skills (e.g., Beck, 2011; Cohen et al., 2006). Many CBT interventions also incorporate a focus on fidelity, such inclusion of the Cognitive Therapy Rating Scale to assess competence and adherence (Beck, 2011). We hypothesized that supervisors with a cognitive-behavioral orientation would show higher odds of engaging in "*directive*" supervision than those who do not endorse a cognitive-behavioral orientation.

In order to provide skillful didactic instruction, accurately model skills, and monitor fidelity to an EBT, supervisors likely need to be knowledgeable about that EBT and believe in their own ability to supervise the EBT. To date, the limited literature that has examined the influence of these constructs on supervision has been mixed. Supervisors' knowledge of a particular EBT did not predict how intensively they focused on EBT-specific content during supervision based on clinician self-report (Lucid et al., 2018) or objectively coded coverage of EBT-related content (Pullmann et al., 2018); however lower levels of EBT knowledge did predict greater non-EBT content coverage (e.g., case background information, administrative work; Pullmann et al., 2018). Supervisor's self-efficacy did not predict their coverage of EBT content (Pullmann et al., 2018). Despite these mixed results, the limited literature warrants further investigation of the relation among supervisor EBT knowledge and self-efficacy and their supervision practices. We hypothesize that greater years of experience supervising, higher EBT knowledge, and increased self-efficacy in supervising a particular EBT will predict higher odds of engaging in "*directive*" supervision.

Clinician characteristics

Clinician characteristics may also influence the practices used in supervision. This may occur indirectly through supervisors reacting to a clinician's characteristics or through the clinician requesting support. Clinician's general experience delivering therapy and their EBT-specific knowledge and experience may influence practices used in supervision. For instance, a supervisor working with a clinician with limited experience delivering therapy and limited experience with a particular EBT may engage in more teaching, provide more clinical suggestions, and model treatment components to support the clinician's EBT delivery. Similarly, a clinician with less treatment experience and EBT-specific experience may ask their supervisor for instruction and modeling on a particular treatment component. Among two studies that have examined the association between clinician characteristics and supervision practices, they found that clinicians' experience conducting therapy and EBT expertise did not predict EBT supervision intensity (Lucid et al., 2018) or EBT content coverage in supervision (Pullmann et al., 2018). Despite the limited support for the influence of clinician characteristics on supervision practice, these constructs warrant further investigation. We hypothesized that fewer years of experience providing therapy, lower objective EBT knowledge, and less experience delivering the EBT with clients in the past 3 months would be associated with greater odds of a supervisor engaging in "*directive*" supervision.

Supervision session characteristics

As noted above, characteristics of a supervision session can also influence what supervisors do during supervision. Client caseloads in routine mental health settings are often large, and estimates of the amount of supervision time per case range from 5 to 12 minutes per case (Choy & Victoria, 2018; Dorsey et al., 2018). High client caseloads may require supervisors to spend more time gathering information about clients than using more

active supervision techniques such as modeling and didactic instruction. The length of supervision sessions can also vary greatly (Choy & Victoria, 2018), and shorter time for supervision may further limit the time available to employ active supervision techniques. We hypothesized that longer supervision session length and the fewer number of clients discussed will be associated with higher odds of engaging in "*directive*" supervision, as this cluster included more intensive technique coverage across all techniques.

Method

Participants

Participants included the 28 supervisors and 70 clinicians who participated in the study reported in Chapter 1.

Measures

Participant Characteristics. Clinicians provided information on the number of years they had conducted therapy, theoretical orientation, the number of children and adolescents they used TF-CBT with in the past three months. Supervisors provided information on years providing supervision and their theoretical orientation. Participant characteristics were measured at baseline.

TF-CBT Knowledge. Supervisors and clinicians completed a 13-item multiple choice TF-CBT knowledge test. This measure was adapted from the Denver Post Health Survey (Fitzgerald, 2010), with items added to reflect those included in the clinician TF-CBT certification program (https://tfcbt.org). Participants completed multiple choice and true-false response ratings to items such as "Unhelpful trauma related thoughts can sometimes be accurate." The measure demonstrated good response range for item difficulty and item discrimination. It has also demonstrated convergent validity with the number of TF-CBT

trainings respondents have attended and TF-CBT self-efficacy (Dorsey, Pullmann, et al., 2017). TF-CBT knowledge was completed at the baseline time point.

Self-Efficacy in Supervision. Supervisors' competence in supervising TF-CBT was measured using the Self-Efficacy in Supervision index. This 13-item index was adapted from a measure created by one of the TF-CBT developers (Deblinger; Child Abuse Research Education and Service Institute, Rowan University, 2013) and the Project BEST team while conducting a statewide implementation of TF-CBT in South Carolina (National Crime Victims Research and Treatment Center, MUSC, 2010). Items were measured on a 5-point Likert scale (1, *not at all*; 2, *a little bit*; 3, *somewhat*; 4, *very much*; 5, *exceptionally*). Sample items include an assessment of supervisors' confidence to: "Supervise clinicians in all of the TF-CBT components," and "Suggest and describe in detail several alternative methods for implementing each TF-CBT component." Self-efficacy was completed at the baseline time point.

Implementation Climate. Supervisors and clinicians completed the Evidence-Based Organizational Checklist to assess the degree to which their organizations expect, support, and reward EBT. Scores from respondents from the same organization were aggregated to create an organizational implementation climate score (Weiner et al., 2011). Item content is similar to that of another implementation climate measure; however that measure was not available at the start of this study (Ehrhart, Aarons, & Farahnak, 2014). The six-item measure is rated on a 4-point Likert scale (1, *never*; 2, *occasionally*; 3, *most of the time*; 4, *ongoing/routinely*). Sample items include, "Executive leadership (e.g., administrators, directors) explicitly and repeatedly express support for and promote use of EBT," and "Clinicians are provided with EBT training opportunities and ready access to EBT materials (manuals, handouts, equipment)." Unidimensionality and internal reliability of measure scores has been supported in previous studies using this measure (Dorsey, Pullmann, et al.,

2017). Internal reliability in the current study was also good (Cronbach's α = 0.86). Higher scores reflect a more supportive EBT implementation climate. The significantly high office-level Intraclass Correlation (ICC (1,1) = 0.41) supports the construct validity of the measure. The ICC is used to reflect "validity" rather than "reliability" because the clustering of implementation climate ratings by members of the same office suggests that perceptions of climate are shared at the office level (Jacobs, Weiner, & Bunger, 2014; Marsh et al., 2012). Due to the small number of supervisors per office, we will include implementation climate in analyses at the supervisor level (e.g., two supervisors in the same office would have the same climate score). EBT implementation climate was completed at the baseline time point.

Supervision Session Time and Cases Discussed. Coders documented the length of the supervision session (in minutes) and number of cases discussed during coding of supervision sessions (described in the method section of Chapter 1).

Supervision style. Supervision sessions were characterized as either "*directive*" or "*undifferentiated low*" based on the results of the cluster analysis conducted in Chapter 1. Supervision sessions that were categorized into the "*directive*" cluster were coded as 1 and those categorized in the "*undifferentiated low*" cluster were coded as 0.

Procedures

Supervisor and clinician study participants completed all self-report measures online upon joining the study. This reflects the baseline timepoint, at the beginning of the descriptive phase and before beginning to record their supervision sessions. Participants received $30 for completing the surveys.

Analytic Plan

A three-level mixed effects logistic regression model with random effects for supervisor and clinician was used to examine predictors of assignment to the "*directive*" cluster identified in Chapter 1. Given that "*directive*" supervision is a dichotomous variable, a

score of zero reflects "*undifferentiated*" supervision. A four-level mixed effects regression model would be more appropriate given that supervisors are also nested within organizations; however, several organizations have a single supervisor, limiting the ability to estimate a four-level model. Models were computed iteratively to test the bivariate association between level-1, level-2, and level-3 predictors and "*directive*" cluster assignment. Slopes were allowed to randomly vary and model fit was evaluated using the overall model-2 Log Likelihood, Aikaike Information Criterion, and Bayesian Information Criterion (-2LL, AIC, and BIC). Allowing slopes to vary did not improve model fit, therefore all slopes were fixed.

Table 7

Independent Variable Correlations[a]

	1	2	3	4	5	6	7	8
1. Number of cases discussed								
2. Supervision session length	0.53							
3. Years providing psychotherapy	-0.04	-0.10						
4. TF-CBT clients seen in past 3 months	-0.06	-0.11	-0.01					
5. Clinician TF-CBT Knowledge	-0.03	-0.05	-0.01	0.04				
6. Supervisor TF-CBT Knowledge	-0.24	-0.22	-0.04	0.22	-0.03			
7. Years providing supervision	-0.07	-0.25	0.20	0.04	0.14	-0.01		
8. Supervision Self-Efficacy	0.02	-0.08	-0.11	0.27	-0.06	0.52	-0.03	
9. EBT Implementation Climate	-0.05	-0.08	-0.03	0.29	0.06	0.24	-0.23	0.35

[a]The categorical variable, supervisor theoretical orientation, is not included in the table.

Results

Predictor means and standard deviations are included in Table 1 and predictor correlations are shown in Table 7. Results for the mixed effects logistic regression model examining predictors of supervision cluster assignment are shown in Table 8. The supervision session-level characteristics, session length and cases discussed, statistically significantly predicted the odds of being assigned to the "*directive*" cluster. For a one-unit increase in the number of cases discussed during a supervision session, the expected odds of being assigned to the "*directive*" supervision cluster decreased by 20% (OR = 0.80; 95% CI = 0.64–0.99) and for each additional minute in the length of a supervision session, there was an expected 11% increase in the odds of being assigned to the "*directive*" cluster (OR = 1.11; 95% CI = 1.07–1.15). The number of TF-CBT clients treated by a clinician in the past three months was statistically significantly associated with decreased odds of being assigned to the "*directive*" cluster (OR = 0.92; 95% CI = 0.87–0.98). For each additional TF-CBT client seen, there was an expected 8% reduction in the odds of being in the "*directive*" cluster. Clinician's knowledge of TF-CBT was marginally significantly related to supervision cluster assignment (OR = 0.85; 95% CI = 0.71–1.01; p = .07). A one-unit increase on a test of TF-CBT knowledge predicted a 15% decrease in the odds of being assigned to the "*directive*" cluster. No supervisor-level predictors were statistically significantly related to supervision cluster assignment.

Table 8

Mixed Effects Logistic Regression Model Predicting Directive Supervision

	B	***SE***	**Exp(B)**	**95% *CI***	***z***	***p***
Intercept	-0.64	0.31	0.53	[0.29, 0.97]	-2.06	0.04
Number of cases discussed	-0.23	0.11	0.80	[0.64, 0.99]	-2.06	0.04
Supervision session length	0.11	0.02	1.11	[1.08, 1.15]	6.42	<0.01
Years providing psychotherapy	0.01	0.03	1.01	[0.95, 1.08]	0.34	0.73
TF-CBT clients seen in past 3 months	-0.08	0.03	0.92	[0.87, 0.98]	-2.64	<0.01
Clinician TF-CBT Knowledge	-0.17	0.09	0.85	[0.71, 1.01]	-1.82	0.07
Supervisor TF-CBT Knowledge	0.22	0.22	1.25	[0.81, 1.93]	1.01	0.31
Years providing supervision	-0.02	0.05	0.98	[0.89, 1.09]	-0.31	0.77
Supervision Self-Efficacy	0.62	0.63	1.86	[0.54, 6.44]	0.98	0.33
Supervisor Theoretical Orientation: Family Systems	1.15	0.82	3.14	[0.63, 15.73]	1.39	0.16
EBT Implementation Climate	0.39	0.31	1.47	[0.81, 2.69]	1.26	0.21
Variance components	**Variance**	**SD**				
Residual	0.90	0.95				
Clinician	0.39	0.63				
Supervisor	1.18	1.08				

Interim Summary

Characteristics of the supervision session and therapist characteristics predicted greater odds of supervisors engaging in "*directive*" supervision. Fewer cases discussed during supervision and longer supervision session duration predicted greater odds of engaging in "*directive*" supervision. When supervising clinicians who had delivered TF-CBT with fewer clients in the past three months, supervisors had greater odds of engaging in "*directive*" supervision. Although marginally significant, when supervising clinicians with less knowledge of TF-CBT, supervisors also showed greater odds of engaging in "*directive*" supervision. No supervisor characteristics statistically significantly predicted the odds of engaging in "*directive*" supervision. At the organization-level, EBT implementation climate was unrelated to engaging in "*directive*" supervision.

The negative association between clinician's experience delivering TF-CBT and "*directive*" supervision provides some support for the validity of the cluster analysis results. Clinicians with greater experience delivering TF-CBT may not require supervision that incorporates higher intensity use of directive techniques such as didactic instruction and experiential learning techniques such as modeling. The marginally significant associations between clinician's TF-CBT knowledge and supervisor's theoretical orientation with "*directive*" supervision were both in the hypothesized direction, providing some, albeit weak, support for the validity of the cluster analysis results. Importantly, there were no results that were in the opposite direction of the hypothesized associations that would challenge the validity of the results. However, there were several hypothesized associations that did not add support for the validity of the cluster analysis results.

CHAPTER 3

HOW SUPERVISION STYLES RELATE TO CLINICIANS' EBT DELIVERY

Introduction

In addition to exploring how workplace-based supervisors are supervising EBT delivery, we sought to understand how the supervision clusters relate to clinicians' EBT delivery. Two studies of workplace-based supervision of EBT have tested EBT-specific models of supervision and found positive effects on clinician adherence and client outcomes. An observational study of Multisystemic Therapy found that supervisor adherence to a manualized MST supervision approach predicted improved clinician adherence (Henggeler et al., 2002) and child outcomes (Schoenwald et al., 2009). Specifically, supervision that attended to clinicians' adherence to MST (i.e., discussions of MST assessment and intervention strategies) predicted greater clinician adherence (Schoenwald et al., 2009). Moreover, supervisors who engaged in greater adherence to the structure and process of MST supervision (e.g., through efficiently managing group supervision) and focused on clinician development (e.g., clinician's goals, skills, and competencies) predicted improved youth behavior and functioning.

Another study compared supervision as usual to a Motivational Interviewing (MI) supervision approach to supervise clinician delivery of MI (Martino et al., 2016). MI supervision included corrective feedback based on review of actual practice and skills coaching using behavioral rehearsal. Clinicians in the MI supervision condition demonstrated greater competency in MI compared to those that received supervision as usual. These studies provide support for the importance of techniques involving practice, observation, feedback, and fidelity monitoring in impacting clinician adherence, competency, and client outcomes.

These studies show that workplace-based supervisors can adopt an EBT-specific supervision approach and have positive effects on clinician EBT delivery and client outcomes. To our knowledge, the current study is the first to examine how workplace-based supervisors

use combinations of techniques in naturally occurring supervision of an EBT and how that relates to EBT delivery. We use the existing evidence for the effect of various supervision techniques on clinician treatment delivery to inform our hypotheses. Although not all five supervision techniques have been empirically investigated, three have received greater support in the literature. Didactic instruction was a supervision technique included in 75% of studies that manipulated supervision and found a positive effect on supervisees' attitudes, emotional self-awareness, motivation, or skills (Milne et al., 2008). As described above, greater supervisor attention to clinician's EBT adherence (i.e., fidelity assessment) during supervision predicted greater clinician EBT adherence during delivery (Schoenwald et al., 2009). In EBT supervision delivered by expert consultants, greater use of modeling predicted higher EBT treatment delivery (Bearman et al., 2013). Although these studies do not speak to the appropriate dosage of these techniques, they have shown a positive association between use of these techniques and treatment delivery. Therefore, we hypothesized that receiving a *higher* proportion of "*directive*" supervision sessions would predict higher fidelity to TF-CBT. Specifically, a higher dose of "*directive*" supervision would be associated with higher intensity delivery of TF-CBT components in the stabilization phase of treatment including psychoeducation, relaxation, affective modulation, and cognitive coping (PRAC). We originally aimed to explore the relation among "*directive*" supervision and fidelity to the trauma narrative. However, due to insufficient power (described below in the method section), we were unable to evaluate this study aim. Alternatively, we examined the relation among "*directive*" supervision and whether or not the trauma narrative, the imaginal exposure component of TF-CBT, was delivered.

Exposure is an element of most well-established treatments for children exposed to traumatic events (Dorsey, McLaughlin, et al., 2017). In TF-CBT, the trauma narrative uses gradual imaginal exposure by creating an individual story about the traumatic event

(Ramirez de Arellano et al., 2014). While some debate exists regarding whether explicit use of exposure, such as the trauma narrative, is essential to achieve improvements in client outcomes, all studies comparing treatments with and without explicit exposure have included some amount of general exposure through other elements such as psychoeducation (Dorsey, McLaughlin, et al., 2017). Clinicians in community mental health often express discomfort with delivering exposure and among children being treated for trauma exposure, exposure was used in less than a quarter of usual care cases (Borntrager, Chorpita, Higa-McMillan, Daleiden, & Starace, 2013). Given this established gap in the delivery of exposure, we examined whether "*directive*" supervision was related to clinicians' delivery of the trauma narrative. We hypothesized there would be a positive relation among the proportion of supervision sessions a clinician received "*directive*" supervision and delivery of the trauma narrative.

Method

Participants

The sample size for participants included in each set of analyses, by outcome and participant type, are described in Table 9. A total of 23 supervisors participated, which included a subset of the supervisors from the previous chapters whose clinicians enrolled a TF-CBT case and submitted at least one TF-CBT session audio recording. Among clinicians who did not submit a tape, some left their organization or did not have an eligible TF-CBT case. A total of 40 clinicians submitted at least one TF-CBT audio file and were included in the current analyses. Among those clinicians, 34 submitted at least one tape that incorporated a stabilization-phase component (i.e., psychoeducation, relaxation, affective modulation, and/or cognitive coping [PRAC]) and 24 submitted at least one tape that incorporated the imaginal exposure component (i.e., trauma narrative [TN]). Among the 60 children included

in Chapter 1, TF-CBT sessions were submitted for 49 of those children by their clinicians. Thirty-nine clients had a PRAC session and 25 of those children had a TN session. Eligibility requirements for participants were as described in Chapter 1.

Table 9

Participant Sample Size by Outcome

	PRAC Fidelity[a]	TN Fidelity[b]	TN Delivery[c]
Participant type			
Supervisor	21	17	23
Clinician	34	24	40
Child	39	25	49

[a]Participants included a subset of those who had at least 1 audio file for the PRAC phase of treatment

[b]Participants included a subset of those who had at least 1 audio file for the trauma narrative

[c]Participants included a subset of those who had at least 1 audio file

Measures

PRAC fidelity. Audio files of clinicians' sessions with clients were behaviorally coded to measure clinician fidelity to TF-CBT. Fidelity was coded using a TF-CBT specific version of the TPOCS developed by the research team. This was adapted from the TF-CBT Checklist Scoring Sheet (Deblinger, unpublished measure, 2005) and Garland's PRAC TPOCS-S (Garland, Bickman, & Chorpita, 2010). Twelve TF-CBT content areas (e.g., relaxation, trauma narrative) and 13 techniques (e.g., teaching, Socratic questioning) were coded for their intensity ranging from 0-6 (0: *nonoccurrence*; 1–2: *low*; 3–4: *medium*; 5–6: *high intensity*). While TF-CBT involves individual child sessions, parent sessions, and conjoint child-parent sessions, the current study uses fidelity scores from individual child sessions to assess PRAC intensity. Two audio files were randomly selected per case, capturing the stabilization phase of TF-CBT treatment involving the psychoeducation, relaxation, affective modulation, and/or cognitive coping components (PRAC) and the explicit exposure phase of treatment involving imaginal exposure through the trauma narrative (TN). For PRAC sessions, fidelity was

calculated using the highest intensity score for any one PRAC content item in the session. For TN sessions, fidelity was calculated using the intensity score for the TN. This yielded a separate fidelity score for the PRAC and TN phases of treatment ranging from 1-6.

Forty eight percent of coded TF-CBT sessions recordings were coded by two or more coders and interrater reliability was computed to ensure every coder maintained reliability standards. Four waves of interrater reliability testing were conducted to protect against coder drift. Using absolute agreement, single measures, and two-way random effects models, ICCs (2,2) were equal to or greater than 0.70 for each of these waves, and averaged 0.76, suggesting excellent reliability (Cicchetti, 1994). At the item-level, ICCs for the PRAC and TN content ranged from 0.77 to 0.84.

Trauma Narrative Delivery. The delivery of the trauma narrative was determined based on the treatment session audio files submitted by the clinician. Each child case either received a 0, indicating they did not receive a trauma narrative, or a 1, indicating they did receive a trauma narrative in the first six months of treatment. The occurrence of the trauma narrative in child or conjoint parent-child sessions was determined using the audio recording labels submitted by clinicians and through listening to the content of the session. Clinicians labeled audio files with the primary session component. For quality assurance, coders would then listen to the audio file to ensure it did include the trauma narrative component. If there was not discussion of the trauma narrative in the audio file, the coder would relabel the session to document the primary session type (i.e., PRAC or another component). If the clinician submitted more than one audio file labeled "trauma narrative," the coder would then listen to this file to assess whether it included the trauma narrative component. If no other sessions were labeled with trauma narrative or if a clinician never submitted an audio file labeled "trauma narrative," they were considered to have not delivered this component.

Dose of "directive" supervision. The proportion of supervision sessions a clinician received "directive supervision", reflecting their dose of "*directive*" supervision, was calculated based on the results from Chapter 1. Proportions ranged from 0 to 1. Given that supervision cluster assignment was dichotomous, the inverse proportion reflects the proportion of sessions a clinician received "*undifferentiated*" supervision (i.e., dose of "*undifferentiated"* supervision).

Procedures

Data collection

Clinicians were asked to audio record all TF-CBT sessions with cases enrolled in the study for a maximum of six months or until treatment was terminated, if that came first. Clinicians labelled the audio recordings with the primary treatment component the session focused on and recordings were date stamped. The TF-CBT audio recordings were saved on study-provided, password-protected tablet devices. Recordings were transferred to the research team using a cloud-based server compliant with the Health Insurance Portability and Accountability Act of 1996.

Coder training

The same six coders who coded the supervision data also coded the TF-CBT fidelity data. Coder training is described in the Chapter 1 method.

Session sampling procedures

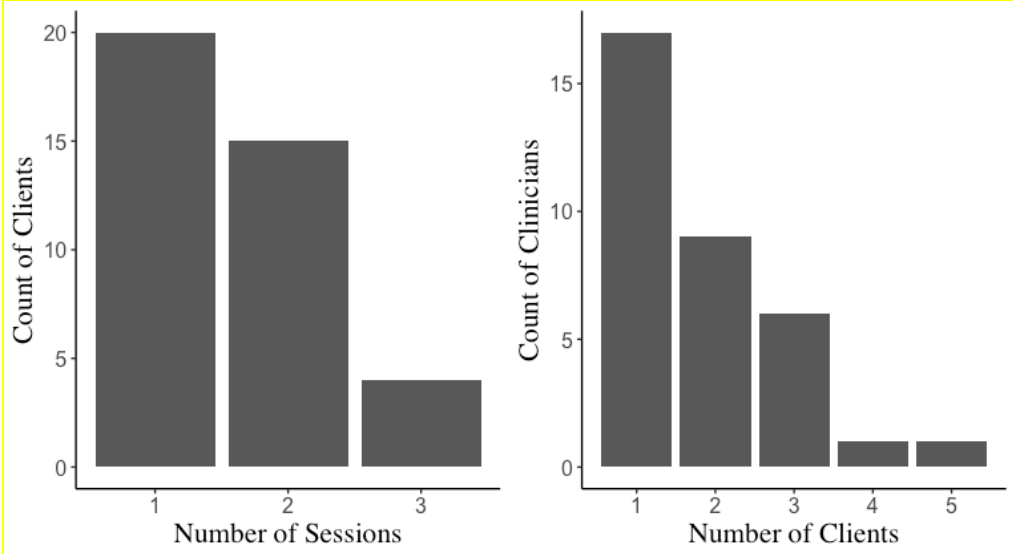
In total, 465 recordings of TF-CBT sessions were received from 40 clinicians with 49 children who participated in the descriptive phase. On average, clinicians submitted 8.61 recordings (*SD* = 4.59; Range = 1-19) during the 6-months any case was enrolled in the study. Two audio files were randomly selected per case, one audio file labeled with a PRAC component and one labeled as the TN. Thirty-nine clients had at least one PRAC session and 25 clients had at least one TN session.

Analytic plan

Directive Supervision Predicting PRAC Fidelity. A three-level mixed-effects regression model with random effects for client and clinician was conducted to examine the relation among the clinicians' dose of "*directive*" supervision and their fidelity to the PRAC components of TF-CBT. While a four-level model accounting for nesting within supervisor would be more appropriate, 50% of supervisors had a single clinician, limiting the ability to estimate a four-level model. The three-level mixed effects model failed to converge and showed no variability in the random effects for client or clinician. A two-level mixed-effects regression model with a random intercept for client was re-estimated and this model also failed to converge, showing no variance in the random effect for the client. The lack of convergence was likely due to high frequency of clinicians and client clusters of n = 1 (see Figure 10), as simulation studies show small cluster sizes often lead to low model convergence rates (Moineddin, Matheson, & Glazier, 2007).

Figure 10

Nesting at the Session and Client Level



The lack of variability at the client and clinician level precluded the calculation of an intra-class correlation (ICC), however this suggested that there may not be meaningful variability to model at the client or clinician level. Given the inability to conduct a mixed effects model, we examined the relation among clinicians' dose of "*directive*" supervision and PRAC fidelity using a generalized estimating equation (GEE) model with an exchangeable correlation structure, accounting for the nesting within client. GEE models account for clustering through using residuals to iteratively estimate a correlation matrix for within-cluster observations which are then used to produce updated estimates of the regression coefficients that account for clustering (McNeish, Stapleton, & Silverman, 2016). We also replicated the analysis using a regression model to compare the results with and without accounting for nesting.

A power analysis revealed that the power to detect an effect of clinicians' dose of "*directive*" supervision on PRAC fidelity with a sample of 39 clients and an anticipated R^2 of 0.16 was 0.77. Previous studies comparing clinician's CBT competence among clinicians who completed CBT training with and without additional supervision showed a large effect size (d= .97; Rakovshik, Mcmanus, Vazquez-montes, Muse, & Ougrin, 2016), while a study comparing bug-in-the-eye to standard supervision on clinician competence found a medium effect size (Weck et al., 2016). Therefore, an anticipated coefficient of determination (R^2) of 0.16 was chosen as this reflects a medium effect size (Cohen, 1992). While a power of 0.77 does not meet the conventional .80 power threshold, we felt this was still high enough to proceed with the analyses.

Dose of "directive" supervision predicting Trauma Narrative Delivery. We initially aimed to examine the relation among clinicians' dose of "*directive*" supervision and clinician fidelity to the imaginal exposure element of TF-CBT, the trauma narrative. However, fewer clinicians submitted sessions that included a trauma narrative, which yielded a sample of 25

clients with a trauma narrative. The power to detect an effect of supervision cluster assignment on fidelity to the trauma narrative with a sample of 25 clients and an anticipated R^2 of 0.16 was 0.67. This is significantly lower than the suggested .80 power, therefore, we did not conduct this model. As an alternative, we examined the relation among clinicians' dose of "*directive*" supervision and whether or not a client received a trauma narrative. The power to detect an effect was 0.78 with a sample of 49 clients. This was calculated under the assumption that a clinician has an estimated probability of .30 for delivering a trauma narrative when they received "*undifferentiated low*" supervision in all supervision sessions and an estimated probability .70 for delivering the trauma narrative when a clinician received "*directive*" supervision in all supervision sessions.

A three-level mixed effects logistic regression model with random effects for clinician and supervisor would be most appropriate to account for the nesting. However, as with the analyses in previous chapters, both a three-level model and two-level model failed to converge and had no variance at the clinician or supervisor level. Given this lack of variability, a logistic regression model was estimated to examine the relation among clinicians' dose of "*directive*" supervision and whether or not the trauma narrative was delivered with a client without accounting for nesting.

Results

Dose of "Directive" Supervision Predicting PRAC Intensity

The average PRAC intensity score, including psychoeducation, relaxation, affective modulation, or cognitive coping, was 4.06 (*SD* = 1.03). The results of the GEE model (Table 10) and the regression model (Table 11) are comparable, suggesting that ignoring nesting did not bias the model estimates. The regression analysis suggested that clinicians' dose of

"*directive*" supervision was not statistically significantly related to their fidelity to the PRAC components of TF-CBT, $B = 0.65$, $t(60) = 1.53$, $p = 0.14$.

Table 10

Generalized Estimating Equations (GEE) Model Predicting PRAC Intensity (n = 62 Sessions; n = 39 Clients; n = 34 Clinicians)

	B	**Robust *SE***	**95% *CI***	**Robust *Z***	***p***
Intercept	3.84	0.21	[3.43, 4.25]	18.28	<.01
Dose of Directive Supervision	0.64	0.38	[-0.11, 1.39]	1.68	0.09

Table 11

Regression Model Predicting PRAC Intensity without Accounting for Nesting (n = 62 Sessions; n = 39 Clients; n = 34 Clinicians)

	B	***SE***	**95 % CI**	***t***	***p***
Intercept	3.85	0.20	[3.45, 4.25]	19.21	<.001
Dose of Directive Supervision	0.65	0.43	[-0.21, 1.1]	1.53	0.14

Notes. $R^2 = 0.04$, $p = 0.14$

Despite the lack of variability in the client-level random effect, the model was re-estimated with a sample of clients contributing a single PRAC session to further explore the influence of nesting on the model results. For clients who had more than one session, a session was randomly sampled to yield a sample of 39 sessions from 39 clients nested within 34 clinicians. A regression model was estimated (Table 12), showing results comparable with the previous GEE and regression analyses suggesting that clinicians' dose of "*directive*" supervision was not statistically significantly related to their fidelity to the PRAC components of TF-CBT, $B = 0.62$, $t(37) = 1.06$, $p = 0.30$.

Table 12

Regression Model Predicting PRAC Intensity (n = 39 Sessions; n = 39 Clients; n = 34 Clinicians)

	B	***SE***	**95 % CI**	***t***	***p***
Intercept	4.00	0.26	[3.28, 4.40]	15.04	<.001
Dose of Directive Supervision	0.62	0.58	[-0.19, 2.26]	1.06	0.30

Notes. R^2 = 0.03, p = 0.29

Dose of "Directive" Supervision Predicting the Trauma Narrative Delivery

Thirty-one clients (63%) received the trauma narrative component of TF-CBT. The results (Table 13) suggest that when clinicians received "*directive*" supervision in all supervision sessions, the odds of a clinician delivering the trauma narrative was 11.85 times higher than a clinician who received "*undifferentiated low*" supervision in all supervision sessions (OR = 12.85; CI = 1.34–123.44). Notably, the confidence interval was quite large, suggesting a statistically significant yet imprecise estimate.

Table 13

Logistic Regression Model Predicting Trauma Narrative Delivery (n = 49 Clients; n = 40 Clinicians)

	B	**Exp(*B*)**	**95 % CI**	***SE***	***z***	***p***
Intercept	-0.24	0.79	[0.33, 1.87]	0.44	-0.54	0.59
Dose of Directive Supervision	2.55	12.85	[1.34, 123.44]	1.15	2.21	0.03

To assess for possible biased estimates due to ignoring the clustering within clinicians, the model was re-estimated using a random sample. For clinicians who had more than one client, a single client was randomly sampled to yield a sample of 40 clients and 40 clinicians. Results were comparable to the original model and suggest as the proportion of supervision sessions a clinician received "*directive*" supervision increased, the odds of a clinician delivering the trauma narrative with their client also increased (OR = 17.24; CI = 1.41–210.84). This model also had a very large confidence interval suggesting a statistically significant yet imprecise estimate.

Table 14

Logistic Regression Model Predicting Trauma Narrative Delivery (n = 40 Clients; n = 40 Clinicians)

	B	**Exp(*B*)**	**95 % CI**	***SE***	***z***	***p***
Intercept	-0.40	0.67	[0.25, 1.79]	0.50	-0.80	0.42
Dose of Directive Supervision	2.85	17.24	[1.41, 210.84]	1.28	2.23	0.03

Interim Summary

The proportion of sessions a clinician received "*directive*" supervision was not significantly related to the intensity with which they delivered PRAC components of TF-CBT, the stabilization phase of treatment including psychoeducation, relaxation, affective modulation, and cognitive coping components. Clinicians who received a higher proportion of "*directive*" supervision were more likely to deliver the trauma narrative with a client, the imaginal exposure component of TF-CBT. However, the confidence interval for this effect was very large, suggesting that this estimate is very imprecise. The true effect can vary from a very meaningful effect to a very insignificant effect of "*directive*" supervision on delivery of the trauma narrative.

The ability to appropriately model the data by accounting for nesting was limited by the very small nested sample sizes at the clinician and supervisor levels. Most supervisors had a single clinician and most clinicians had a single case nested under them. The limited variability in the random intercept estimates suggest that the effect of nesting may have been negligible, and steps were taken to replicate the analysis using multiple analytic techniques and by taking a random sample of the data to eliminate nesting and then reanalyzing the data. The consistency of the results across these methods provides some additional support for the non-significant relation among "*directive*" supervision and PRAC intensity. However, we were only powered to detect a medium effect size. Therefore, it is possible that there may have been a smaller effect of "*directive*" supervision on PRAC that we were not powered to detect. Additionally, the limited number of clients who had a session focused on the trauma narrative reduced the power to detect an effect of "*directive*" supervision on fidelity to the trauma narrative, therefore, the original study aim was modified.

CHAPTER 4
DISCUSSION OF MAJOR FINDINGS

Discussion

This study aimed to contribute to the limited knowledge of supervision methods used in workplace-based supervision to support clinicians delivering an EBT. We advance the limited characterizations of workplace-based supervision by taking a cluster approach to characterize subgroups of supervision styles used by workplace-based supervisors. The results provide insight into supervision session characteristics and clinician experience that predict two styles of supervision used and how those supervision styles relate to clinicians' delivery of EBT components.

The results revealed two supervision session clusters, termed "*directive*" and "*undifferentiated low*" supervision. There were several techniques that did not reliably distinguish between supervision sessions and were not represented in the clusters. Rather than reflecting a particular "supervision style," our findings suggest that supervisors tended to tailor their supervision to the clinician they were supervising. The odds of a supervision session being "*directive*" were higher when the supervision session was longer, involved discussion of fewer clients, and the clinician had less experience delivering TF-CBT in the past three months. Clinicians who received a higher proportion of "*directive*" supervision sessions had greater odds of delivering the trauma narrative, the imaginal exposure element of TF-CBT, with a client.

The original goal of the cluster analysis was to examine subgroups of supervision across eleven supervision techniques. However, six techniques ultimately were excluded from the cluster analysis using an iterative exclusion approach that was informed by the theoretical importance of a particular technique and its performance in analytically distinguishing between clusters. Among the excluded techniques, three were considered to be 'gold standard' supervision techniques, as they are typically included in efficacy trials to ensure high-quality treatment delivery (Roth et al., 2010) and found to support training

outcomes in EBT training efforts (Beidas & Kendall, 2010; Herschell et al., 2010). These three techniques included behavioral rehearsal, review of practice, and symptom monitoring. While these 'gold standard' techniques hold theoretical significance, the inclusion of unnecessary or non-informative variables has been show to degrade results of cluster-based analysis (Raftery & Dean, 2006). Reducing the techniques to those that were most discriminative in our sample does pose the risk of overfitting the clusters to our data, limiting generalizability. However, similarities in the use of supervision techniques in the current study and a similar study describing workplace-based supervision (Bailin et al., 2018) provides support for the generalizability of our findings. Additionally, the convergent validity of the resulting clusters was also supported through the significant associations with clinician's TF-CBT experience and trauma narrative delivery.

The supervision clusters showed important distinctions in the use of techniques that have been studied in the literature. Didactic instruction was the highest or second highest intensively used technique in "*undifferentiated low*" and "*directive*" supervision respectively. While didactic instruction is a common technique used in training (Beidas & Kendall, 2010), supervision (Milne et al., 2008), and consultation (Edmunds, Kendall, et al., 2013), studies evaluating training efforts demonstrate that didactic instruction without experiential learning techniques is not sufficient to impact clinician's behavior and client improvement (Beidas & Kendall, 2010). "*Directive*" supervision incorporated experiential learning through the use of modeling, but this was notably absent in "*undifferentiated low*" supervision. Both clusters include low intensity fidelity assessment, although "*directive*" supervision included over a one-point higher average intensity. It is not entirely clear how meaningful a one-point difference in fidelity assessment may be in impacting clinician's EBT delivery, as most studies that examine this association do not provide descriptions of the dosage of these techniques.

One study found a 12% increase in therapist EBT adherence when comparing supervisors with the highest and lowest focus on adherence in supervision (Schoenwald et al., 2009).

Empirical research on clinical suggestions and elicitation is lacking, however, theoretical perspectives provide some insight into how supervisors can use these techniques to support clinician learning. "*Directive*" supervision included moderately intensive clinical suggestions and low intensive elicitation, whereas "*undifferentiated low*" supervision included low intensive clinical suggestions and essentially no elicitation. Theory suggests that effective cognitive-behavioral supervision can support clinician learning through the use of scaffolding strategies (James et al., 2008), specifically through the use of asking questions. In the current study, elicitation involved a supervisor asking questions to a) encourage and elicit clinician thinking and planning for a subsequent session and/or b) to help the clinician to evaluate their own effectiveness in a previous session. James and colleagues (2008) suggest questioning can be used to help the supervisor understand the clinician's knowledge and also to facilitate the clinician's learning by eliciting their own ideas and providing feedback or clinical suggestions. Our findings suggested that clinicians with less experience in delivering TF-CBT and, though only marginally significant, less TF-CBT knowledge were more likely to receive "*directive*" supervision. That is, clinicians who had delivered TF-CBT with one fewer client had 8% higher odds of receiving directive supervision and those who scored one-point lower on the 13-item knowledge test had 15% higher odds of receiving "*directive"* supervision. Observing that these clinicians had less TF-CBT experience and declarative knowledge, supervisors may have relied more heavily on directing and modeling, while incorporating some elicitation to drive clinician learning forward.

Although supportive listening and information gathering did not discriminate between clusters, on average they were both used with moderate intensity across the entire sample. Supportive listening included reflections, validation, acknowledgement and praise

for the clinician. Bailin and colleagues (2018) similarly found frequent use of praise and empathy during supervision sessions. A high level of perceived supervisor support is suggested to be necessary for effective supervision (Russell & Petrie, 1994). The high degree of supportive listening across *both* clusters suggests that even when supervisors incorporate more intensive use of directive techniques in "*directive*" supervision, they are still able to maintain a supportive supervision environment.

While there is a growing body of experimental and correlational research examining the effect of supervision practice on clinical practice (e.g., Alfonsson, Parling, Spännargård, Andersson, & Lundgren, 2018; Martino et al., 2016; Milne et al., 2008; Schoenwald et al., 2009), few studies have examined what factors impact supervision practices. Of those that exist, several rely fully or in part on the sample of supervisors and clinicians included in the current study (e.g., Dorsey, Pullmann, et al., 2017; Lucid et al., 2018; Pullmann et al., 2018). Among this literature, the studies tend to also show null findings with regard to the relation among clinician and supervisor characteristics and supervision practices, though they contrast our findings with regard to clinician experience and EBT implementation climate.

With regard to clinician characteristics, our findings suggest that clinician's experience delivering TF-CBT was related to their receipt of "*directive*" supervision, but their general years of experience delivering therapy was not. Our findings also suggest that supervisors tended to tailor supervision to the clinician. In line with our results, other studies found no support for a relation among clinician's years providing therapy and clinician-reported EBT supervision intensity (Lucid et al., 2018), coded EBT content coverage in supervision (Pullmann et al., 2018), or clinician-reported time allocated to therapy interventions and case conceptualization during supervision (Dorsey, Pullmann, et al., 2017). In contrast to our results, these three studies also found no support for a relation between clinician's TF-CBT experience.

The lack of support for a relation among any supervisor characteristics and their use of "*directive*" supervision, though inconsistent with our hypotheses, largely aligns with the limited existing literature. There is some mixed evidence regarding whether supervision practices are related to supervisors' theoretical orientation and their knowledge of an EBT. Pullmann and colleagues (2018) found that supervisors with a CBT orientation covered assessment with greater intensity than those with a family systems orientation, though this was non-significant after accounting for the influence of EBT implementation climate. Bailin and colleagues (2018) did not find an association between supervisor orientation and the use, duration, or competency of objectively coded EBT content coverage or evidence-based supervision techniques. Supervisors' knowledge of a particular EBT did not predict how intensively they focused on EBT-specific content during supervision based on clinician self-report (Lucid et al., 2018) or objectively coded coverage of EBT-related content (Pullmann et al., 2018). Among the other supervisor characteristics examined, supervisors' years of experience supervising and their perceived self-efficacy supervising TF-CBT did not relate to supervision time allotted to EBT relevant activities (Dorsey, Pullmann, et al., 2017) or the intensity with which they supervised particular EBT content (Lucid et al., 2018). This study provides further support that supervisors' general experience and beliefs about their ability to supervise a particular EBT does not predict the amount of time or intensity with which they supervise an EBT.

The limited evidence for the importance of supervisor characteristics may be explained, in part, by the homogeneity in the supervisors included in these studies. Our sample of supervisors – and that of the studies that rely on the same sample (Dorsey, Pullmann, et al., 2017; Lucid et al., 2018; Pullmann et al., 2018) – is somewhat homogenous, with seventy-five percent of supervisors endorsing a cognitive-behavioral orientation. In a sample of supervisors from varied theoretical backgrounds, supervisors' theoretical

orientation influenced how their supervisees perceived them to be supervising (Putney, Worthington, & McCullough, 1992). When supervisors endorsed a humanistic-psychodynamic orientation, their supervisees perceived them to adhere more closely to a relationship model of supervision relative to supervisors endorsing a cognitive-behavioral orientation.

Interestingly, EBT implementation climate was the most consistent predictor of supervision practices in past studies, yet it did not predict supervision cluster assignment in the current study. Specifically, having a more positive EBT implementation climate predicted greater clinician report of time allotted to EBT activities in supervision (Dorsey et al., 2017), greater clinician-reported EBT supervision intensity (Lucid et al., 2018) and greater objectively coded intensity of coverage of EBT content elements (Pullmann et al., 2018). This discrepant finding, along with the fact that these studies found no clinician characteristics or supervisor characteristics that predicted supervision practices, may reflect an important distinction between the supervision practices examined in these studies and the current study. The supervision practices in these studies include some reference to the content discussed in supervision (Dorsey, Pullmann, et al., 2017b; Lucid et al., 2018; Pullmann et al., 2018), whereas the current study only focuses on techniques used. If an EBT is expected, rewarded and supported in an organization, in order to supervise that EBT, there needs to be discussion of the content related to that EBT. Whereas, the techniques used to support an EBT may be agnostic to the specific EBT. In an organization where EBT is expected, rewarded and supported, supervisors may be selecting the techniques that are best suited to the experience of the clinician.

Supervision practices are also influenced by the characteristics of a particular supervision session. In the descriptive study characterizing the intensity of supervision techniques across the entire sample, Dorsey and colleagues (2018) found that more variance

in the intensity of techniques used in supervision was attributable to the supervision session relative to that attributable to the clinician or supervisor. When there was more time for supervision and fewer cases to review, supervisors were more likely to engage in "*directive*" supervision. Clinicians in routine mental health settings often carry large client caseloads, and estimates suggest the time to discuss each case ranges from 5 to 12 minutes (Choy & Victoria, 2018; Dorsey et al., 2018). Most clinicians and supervisors endorse wanting more time in supervision to spend on functions that are most relevant to EBT (Dorsey, Pullmann, et al., 2017). Exploring creative ways to increase time and decrease the number of cases to discuss in supervision may allow supervisors to engage in more directive supervision when warranted.

While receiving a higher proportion of "*directive*" supervision did not impact the quality with which clinicians delivered components from the stabilization phase of TF-CBT (i.e., psychoeducation, relaxation, affective modulation, and cognitive coping components), it did impact whether or not they delivered imaginal exposure. Studies that have manipulated the techniques used in workplace-based supervision have established a link between supervision and clinician treatment adherence (Henggeler et al., 2002) and treatment competency (Martino et al., 2016). However, some of these manipulated techniques were largely absent in the naturally occurring supervision in the present study. Martino and colleagues (2016) had supervisors provide corrective feedback based on review of actual practice and skills coaching using behavioral rehearsal. In our study, review of actual practice and behavioral rehearsal were nearly non-existent across the two supervision clusters. Schoenwald and colleagues (2009), found that supervision that attended to clinicians' adherence to Multisystemic Therapy (MST) through discussions of MST assessment and intervention strategies predicted greater clinician adherence to MST. "*Directive*" supervision

included higher intensity use of fidelity assessment, though it was still in the low-to-moderate range. This dose of fidelity assessment may not be high enough to impact PRAC intensity.

Alternatively, PRAC skills may be delivered with high intensity regardless of supervision characteristics. Most clinicians in this sample endorsed a CBT theoretical orientation, and relaxation and other techniques focused on developing coping skills are common components in cognitive behavioral treatments for traumatic stress and anxiety (Chorpita & Daleiden, 2009). Clinicians delivering cognitive behavioral treatments often express a preference for treatment components focused on coping skills, such as progressive muscle relaxation or deep breathing, over exposure-based components (Chu et al., 2015; Edmunds et al., 2014). Clinicians' strong preference for and familiarity with treatment components that develop coping skills may diminish the importance of supervision in developing competency in these components.

The greater odds of clinicians delivering the trauma narrative when they received a higher proportion of "*directive*" supervision may reflect an impact on clinician's comfort with delivering exposure. Adoption of exposure significantly lags behind other evidence-based practice elements for youth (Higa-McMillan, Francis, Rith-Najarian, & Chorpita, 2016; Higa-McMillan, Kotte, Jackson, & Daleiden, 2017) and community-based clinicians have rated it as the most difficult strategy to implement among CBT strategies (Chu et al., 2015). Community mental health clinicians often express discomfort with delivering exposure, using it with less than a quarter of children suffering from trauma exposure in usual care settings (Borntrager et al., 2013). The incorporation of modeling in "*directive*" supervision may have improved clinician's comfort with delivering imaginal exposure. Alternatively, clinicians who received a higher dose of "*directive*" supervision may have felt delivering the trauma narrative was expected of them, regardless of their comfort delivering it.

The influence of "*directive*" supervision on delivery of the trauma narrative may also reflect an impact on clinician's pacing of treatment delivery. Delivery of the trauma narrative was measured in the first six months of delivering TF-CBT. The entire treatment can typically be delivered in around 12 – 16 sessions (Cohen et al., 2006). However, treatment duration in community mental health tends to be much longer, with average duration estimates of about 23 sessions (Garland, Brookman-Frazee, et al., 2010; Hawley & Weisz, 2005), which, if delivered weekly, would be about 6 months. This longer treatment duration and slower treatment pacing may result from the differences observed between effectiveness trials and community mental health, which in addition include lower client attendance, increased client complexity, or drift from the intervention (Depp & Lebowitz, 2007). The "*directive*" supervision's higher intensity inclusion of fidelity assessment, which includes discussion of treatment pacing, keeping the treatment brief, and the client's ability to move on to the next component, may have facilitated more timely pacing through the treatment model, allowing the clinician to deliver the trauma narrative within the six month window.

Strengths and Limitations

This study had a number of strengths. First, this study used behaviorally coded supervision sessions and behaviorally coded TF-CBT sessions, providing high-quality measures of supervision techniques and TF-CBT fidelity. The use of behaviorally coded measures also eliminates common method variance that is present in studies that rely solely on self-report measures. Among the existing literature on supervision, much of it has examined supervision in the context of studies utilizing external Doctoral-level supervisors or consultants. This study focused on workplace-based supervision and delivery of treatment in community mental health settings, by clinicians employed in these settings, increasing the

external validity of the study findings. The use of a cluster approach allowed for the characterization of supervision across a combination of various supervision techniques.

A number of limitations of the current study should also be noted. First, the supervision sessions reflect supervision focused on TF-CBT and may not reflect supervision of other EBTs or non-EBTs. Second, the sample was drawn from organizations that participated in a state-funded EBT initiative, which may limit the generalizability to supervisors and clinicians in organizations that have not participated in an EBT training initiative. Additionally, this may have contributed to the homogeneity of the supervisors who participated in the study, as they were interested in supervising a CBT intervention.

The cluster analysis was based on nested data and this analytic technique does not account for this clustering, possibly biasing the clustering results. Moreover, the results of our original cluster analysis did not produce reliable and valid results. We took a theoretical and data-driven approach to reduce the supervision techniques to those that would reliably distinguish between supervision clusters. This approach may have reduced the generalizability of the results. Additionally, the sample size included in the Chapter 3 analyses was small and the limited nesting led to challenges with estimating random effects. The Chapter 3 analyses were also underpowered, requiring the use of an alternative study aim. Though we found a significant effect of "*directive"* supervision on trauma narrative delivery, the confidence interval was very large, providing uncertainty about the practical significance of this result. Client severity was not accounted for in this model which may have impacted whether or not the clinician delivered the trauma narrative. Finally, clients were only followed for six months, so it is possible that the trauma narrative was eventually delivered after that period.

Conclusions

To our knowledge, this is the first study to examine the presence of subgroups of supervision based on objectively coded supervision techniques used to support clinician EBT delivery in community mental health. This study identified two types of supervision, "*directive*" and "*undifferentiated low*" supervision, and their use had an influence on whether or not clients received a trauma narrative, an important component of TF-CBT. The results suggest supervision session characteristics, including supervision time and number of clients discussed, that could be targeted to increase supervisor's ability to use "*directive*" supervision. It also suggests that supervisors are tailoring their supervision style based on clinician's experience with the intervention, reflecting a skillful augmentation of supervision.

References

Alfonsson, S., Parling, T., Spännargård, Å., Andersson, G., & Lundgren, T. (2018). The effects of clinical supervision on supervisees and patients in cognitive behavioral therapy: a systematic review. *Cognitive Behaviour Therapy*, *47*(3), 206–228. https://doi.org/10.1080/16506073.2017.1369559

Bailin, A., Bearman, S. K., & Sale, R. (2018). Clinical supervision of mental health professionals serving youth: Format and microskills. *Administration and Policy in Mental Health and Mental Health Services Research*, *45*(5), 800–812. https://doi.org/10.1007/s10488-018-0865-y

Banerjee, M., Capozzoli, M., Mcsweeney, L., & Sinha, D. (1999). Beyond kappa: A review of measures of interrater agreement measures. *The Canadian Journal of Statistics*, *27*(1), 3–23. https://doi.org/10.2307/3315487

Bearman, S. K., Schneiderman, R. L., & Zoloth, E. (2016). 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

Beck, J. S. (2011). *Cognitive Behavior Therapy: Basics and Beyond* (Second). New York: The Guilford Press.

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

Bennett-Levy, J., McManus, F., Westling, B. E., & Fennell, M. (2009). Acquiring and refining CBT skills and competencies: Which training methods are perceived to be most effective? *Behavioural and Cognitive Psychotherapy*, *37*(5), 571–583. https://doi.org/10.1017/S1352465809990270

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

Brookman-Frazee, L., Haine, R. A., Baker-Ericzén, M., Zoffness, R., & Garland, A. F. (2010). Factors associated with use of evidence-based practice strategies in usual care youth psychotherapy. *Administration and Policy in Mental Health and Mental Health Services Research*, *37*(3), 254–269. https://doi.org/10.1007/s10488-009-0244-9

Chorpita, B. F., & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting*, *77*(3), 566–579. https://doi.org/10.1037/a0014565

Choy, M., & Victoria, B. (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

Chu, B. C., Crocco, S. T., Arnold, C. C., Brown, R., Southam-Gerow, M. A., & Weisz, J. R. (2015). Sustained implementation of cognitive-behavioral therapy for youth anxiety and depression: Long-term effects of structured training and consultation on therapist practice in the field. *Professional Psychology: Research and Practice*, *46*(1), 70–79. https://doi.org/10.1037/a0038000

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

Cicchetti, D. V, & Feinstien, a R. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology*, *43*(6), 551–558. https://doi.org/10.1016/0895-4356(90)90159-M

Cohen, J. (1992). A power primer. *Quantitative Methods in Psychology*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cohen, J. A., Mannarino, A. P., & Deblinger, E. (2006). *Treating trauma and traumatic grief in children and adolescents*. New York: Guilford Publications.

Depp, C., & Lebowitz, B. D. (2007). Clinical trials: Bridging the gap between efficacy and effectiveness. *International Review of Psychiatry*, *19*(5), 531–539. https://doi.org/10.1080/09540260701563320

Dorsey, S., Kerns, S. E. U., Lucid, L., Pullmann, M. D., Harrison, J. P., Berliner, L., … Deblinger, E. (2018). Objective coding of content and techniques in workplace-based supervision of an EBT in public mental health. *Implementation Science*, *13*(1), 1–12. https://doi.org/10.1186/s13012-017-0708-3

Dorsey, S., McLaughlin, K. A., Kerns, S. E. U., Harrison, J. P., Lambert, H. K., Briggs, E. C., … 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*, *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., … Garland, A. F. (2013). Improving practice in community-based settings: A randomized trial of supervision - study protocol. *Implementation Science : IS*, *8*, 89. 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*, *0*(0), 1–15. https://doi.org/10.1007/s10488-017-0796-z

Edmunds, J. M., Beidas, R. S., & Kendall, P. C. (2013a). Dissemination and implementation of evidence – based practices: Training and consultation as implementation strategies. *Clinical Psychology: Science and Practice*, *20*(2), 152–165.

Edmunds, J. M., Beidas, R. S., & Kendall, P. C. (2013b). Dissemination and implementation of evidence – based practices: Training and consultation as implementation strategies. *Clinical Psychology: Science and Practice*, *20*(2), 152–165.

Edmunds, J. M., Kendall, P. C., Ringle, V. A., Read, K. L., Brodman, D. M., Pimentel, S. S., & Beidas, R. S. (2013). An examination of behavioral rehearsal during consultation as a predictor of training outcomes. *Administration and Policy in Mental Health and Mental Health Services Research*, *40*(6), 456–466. https://doi.org/10.1007/s10488-013-0490-8

Ehrhart, M. G., Aarons, G. A., & Farahnak, L. R. (2014). Assessing the organizational context for EBP implementation: The development and validity testing of the Implementation Climate Scale (ICS). *Implementation Science*, *9*(1), 157. https://doi.org/10.1186/s13012-014-0157-1

Everitt, B., Landau, S., & Leese, M. (2001). *Cluster Analysis* (4th ed.). London: Arnold.

Falender, C. A., Cornish, J. A. E., Goodyear, R., Hatcher, R., Kaslow, N. J., Leventhal, G., … Grus, C. (2004). Defining competencies in psychology supervision: A consensus statement. *Journal of Clinical Psychology*, *60*(7), 771–785. https://doi.org/10.1002/jclp.20013

Garland, A. F., Bickman, L., & Chorpita, B. F. (2010). Change what? Identifying quality improvement targets by investigating usual mental health care. *Administration and Policy in Mental Health and Mental Health Services Research*, *37*(1–2), 15–26. https://doi.org/10.1007/s10488-010-0279-y

Garland, A. F., Brookman-Frazee, L., Hurlburt, M. S., Accurso, E. C., Zoffness, R. J., Haine-Schlagel, R., & Ganger, W. (2010). Mental health care for children with disruptive behavior problems: A view inside therapists' offices. *Psychiatric Services*, *61*(8), 788–795. https://doi.org/10.1176/appi.ps.61.8.788

Garland, A. F., Kruse, M., & Aarons, G. A. (2003). Clinicians and outcome measurement: What's the use? *Journal of Behavioral Health Services & Research*, *30*(4), 393–405.

Hawley, K. M., & Weisz, J. R. (2005). Youth versus parent working alliance in usual clinical care: Distinctive associations with retention, satisfaction, and treatment outcome. *Journal of Clinical Child and Adolescent Psychology*, *34*(1), 117–128. https://doi.org/10.1207/s15374424jccp3401

Henggeler, S. W., Schoenwald, S. K., Liao, J. G., Letourneau, E. J., Daniel, L., Henggeler, S. W., … Letourneau, E. J. (2002). Transporting efficacious treatments to field settings: The link between supervisory practices and therapist fidelity in MST programs. *Journal of Clinical Child & Adolescent Psychology*, *31*(2), 155–167. https://doi.org/10.1207/S15374424JCCP3102

Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics and Data Analysis*, *52*(1), 258–271. https://doi.org/10.1016/j.csda.2006.11.025

Herschell, A. D., Kolko, D. J., Baumann, B. L., & Davis, A. C. (2010). The role of therapist training in the implementation of psychosocial treatments: A review and critique with recommendations. *Clinical Psychology Review*, *30*(4), 448–466. https://doi.org/10.1016/j.cpr.2010.02.005

Higa-McMillan, C. K., Francis, S. E., Rith-Najarian, L., & Chorpita, B. F. (2016). Evidence base update: 50 years of research on treatment for child and adolescent anxiety. *Journal of Clinical Child and Adolescent Psychology*, *45*(2), 91–113. https://doi.org/10.1080/15374416.2015.1046177

Higa-McMillan, C. K., Kotte, A., Jackson, D., & Daleiden, E. L. (2017). Overlapping and non-overlapping practices in usual and evidence-based care for youth anxiety. *Journal of Behavioral Health Services and Research*, *44*(4), 684–694.

https://doi.org/10.1007/s11414-016-9502-2

Hogarty, K. Y., Kromrey, J. D., Ferron, J. M., & Hines, C. V. (2004). Selection of variables in exploratory factor analysis: An empirical comparison of a stepwise and traditional approach. *Psychometrika*, *69*(4), 593–611. https://doi.org/10.1007/BF02289857

Hsu, M. L. A., & Fan, H.-L. (2010). Organizational innovation climate and creative outcomes: Exploring the moderating effect of time pressure. *CREATIVITY RESEARCH JOURNAL*, *22*(4), 378–386. https://doi.org/10.1037/a0023537

Hutchinson, T. P. (1993). Kappa muddles together two sources of disagreement: Tetrachoric correlation is preferable. *Research in Nursing & Health*, *16*(4), 313–316.

Jacobs, S. R., Weiner, B. J., & Bunger, A. C. (2014). Context matters: measuring implementation climate among individuals and groups. *Implementation Science*, *9*(1), 46. https://doi.org/10.1186/1748-5908-9-46

James, I. A., Milne, D. L., & Morse, R. (2008). Microskills of clinical supervision: Scaffolding skills. *Journal of Cognitive Psychotherapy*, *22*(1), 29–36. https://doi.org/10.1891/0889.8391.22.1.29

Jurman, G., Riccadonna, S., Visintainer, R., Furlanello, C., & Fondazione, B. K. (2009). Canberra distance on ranked lists. In *Proceedings, Advances in Ranking – NIPS 09 Workshop* (pp. 22–27). Retrieved from https://pdfs.semanticscholar.org/a298/02f53f3bcf348e2c9a95df748277ae7571c5.pdf

Kilminster, S. M., & Jolly, B. (2000). Effective supervision in clinical practice settings: a literature review. *Medical Education (MED EDUC)*, *34*, 827–840.

Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2006). Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science and Practice*, *10*(3), 288–301. https://doi.org/10.1093/clipsy.bpg025

Lucid, L., Meza, R., Pullmann, M. D., Jungbluth, N., Deblinger, E., & Dorsey, S. (2018). Supervision in community mental health: Understanding intensity of EBT focus. *Behavior Therapy*, *49*(4), 481–493. https://doi.org/10.1016/j.beth.2017.12.007

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom Climate and Contextual Effects : Conceptual and Methodological Issues in the Evaluation of Group-Level Effects Classroom Climate and Contextual Effects : Conceptual and Methodological Issues in the Evaluation of Group-Level Effects. *Educational Psychologist*, *47*(2), 106–124. https://doi.org/10.1080/00461520.2012.670488

Martino, S., Paris, M., Añez, L., Nich, C., Canning-Ball, M., Hunkele, K., … 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., & Weisz, J. R. (2010). The Therapy Process Observational Coding System for Child Psychotherapy-Strategies Scale. *Journal of Clinical Child and Adolescent Psychology : The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53*, *39*(3), 436–443. https://doi.org/10.1080/15374411003691750

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2016). On the Unnecessary Ubiquity of Hierarchical Linear Modeling. *Psychological Methods*, *58*(12), 7250–7257. https://doi.org/10.1037/met0000078

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. L. (2009). *Evidence-Based Clinical Supervision: Principles and Practice*. *John Wiley & Sons*.

Milne, D. L., Aylott, H., Fitzpatrick, H., & Ellis, M. V. (2008). How does clinical supervision work? Using a "best evidence synthesis" approach to construct a basic model of supervision. *The Clinical Supervisor*, *27*(2), 170–190. https://doi.org/10.1080/07325220802487915

Milne, D. L., & James, I. (2000). A systematic review of effective cognitive-behavioural supervision. *British Journal OfClinical Psychology*, *39*, 111–127.

Milne, D. L., Reiser, R. P., Cliffe, T., Breese, L., Boon, A., Raine, R., & Scarratt, P. (2011). A qualitative comparison of cognitive-behavioural and evidence-based clinical supervision. *Cognitive Behaviour Therapist*, *4*(4), 152–166. https://doi.org/10.1017/S1754470X11000092

Moineddin, R., Matheson, F. I., & Glazier, R. H. (2007). A simulation study of sample size for multilevel logistic regression models. *BMC Medical Research Methodology*, *7*, 1–10. https://doi.org/10.1186/1471-2288-7-34

Nakamura, B. J., Selbo-Bruns, A., Okamura, K., Chang, J., Slavin, L., & Shimabukuro, S. (2014). Developing a systematic evaluation approach for training programs within a train-the-trainer model for youth cognitive behavior therapy. *Behaviour Research and Therapy*, *53*(1), 10–19. https://doi.org/10.1016/j.brat.2013.12.001

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, *75*(6), 829–841. https://doi.org/10.1037/0022-006X.75.6.829

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*(October). https://doi.org/10.3389/fpubh.2018.00280

Putney, M. W., Worthington, E. L., & McCullough, M. E. (1992). Effects of Supervisor and Supervisee Theoretical Orientation and Supervisor-Supervisee Matching on Interns' Perceptions of Supervision. *Journal of Counseling Psychology*, *39*(2), 258–265. https://doi.org/10.1037/0022-0167.39.2.258

Raftery, A. E., & Dean, N. (2006). Variable selection for model-based clustering. *Journal of the American Statistical Association*, *101*(473), 168–178. https://doi.org/10.1198/016214506000000113

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, *84*(3), 191–199.

Ramirez de Arellano, M., Russell, L., Jobe-Shields, L., George, P., Dougherty, R., Daniels, A., …

Delphin-Rittmon, M. (2014). Trauma focused CBT: assessing the evidence. *Psychiatr Serv.*, *65*(5), 591–602. https://doi.org/10.1176/appi.ps.201300255.Trauma-Focused

Reiser, R. P., & Milne, D. L. (2017). *A Manual for Evidence-Based CBT Supervision*.

Roth, A. D., Pilling, S., & Turner, J. (2010). Therapist Training and Supervision in Clinical Trials : Implications for Clinical Practice, (April), 291–302. https://doi.org/10.1017/S1352465810000068

Russell, R. K., & Petrie, T. (1994). Issues in training effective supervisors. *Applied and Preventive Psychology*, *3*(1), 27–42. https://doi.org/10.1016/S0962-1849(05)80106-4

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

Shalizi, C. (2009). *Distances between Clustering , Hierarchical Clustering*. *Data Mining*. Retrieved from www.stat.cmu.edu/~cshalizi/350

Webster-Stratton, C. H., Reid, M. J., & Marsenich, L. (2014). Improving Therapist Fidelity During Implementation of Evidence-Based Practices: Incredible Years Program. *Psychiatric Services*, *65*(6), 789–795. https://doi.org/10.1176/appi.ps.201200177

Weck, F., Jakob, M., Neng, J. M. B., Höfling, V., Grikscheit, F., & Bohus, M. (2016). The Effects of Bug-in-the-Eye Supervision on Therapeutic Alliance and Therapist Competence in Cognitive-Behavioural Therapy: A Randomized Controlled Trial. *Clinical Psychology and Psychotherapy*, *23*(5), 386–396. https://doi.org/10.1002/cpp.1968

Weiner, B. J., Belden, C. M., Bergmire, D. M., & Johnston, M. (2011). The meaning and measurement of implementation climate. *Implementation Science*, *6*(1), 78. https://doi.org/10.1186/1748-5908-6-78