2. Specific Aims

Effect sizes for evidence-based practices (EBPs) in efficacy trials for child and adolescent mental health are substantial, showing improvements in functioning and well being. In contrast, effect sizes for treatment delivered in "real-world" settings (i.e., usual care) are near zero (Weisz et al., 2006). Efforts to improve outcomes in real-world settings by training clinicians in EBPs have consistently pointed to the importance of focusing on factors that support implementation (Fixsen et al., 2005). In two recent reviews of clinician training, supervision was highlighted as: (1) a critical factor, particularly when sustainability is considered; and (2) a factor that has received limited research attention (Beidas & Kendall, 2010; Herschell et al., 2010). Supervision is commonly provided in most community-based settings (Schoenwald et al., 2008). However, very limited research, most non-experimental, has focused on how to utilize existing, community based supervisors to support clinicians in providing EBPs (Schoenwald et al., 2009). Experimental investigations of supervision strategies that are effective and feasible for improving clinician practice in "real world" settings are needed to reduce the gap between routine and ideal care and improve child outcomes.

In efficacy trials, supervision often includes "gold standard" strategies to increase clinician fidelity (Milne, 2009; Sheidow et al. 2008). These include fidelity monitoring, regular assessment of client symptoms, observation of practice, and ongoing clinician skill-building. Supervision approaches employing combinations of these strategies have been associated with increased fidelity, and fidelity, in turn, has been shown to be associated with more positive client outcomes (Herschell et al. 2010; Schoenwald et al., 2009). When supervision in community-based settings is considered, the limited available research, much of which was conducted by our team, suggests that gold standard strategies are infrequently or inconsistently used (Accurso, Taylor, & Garland, 2011; Carroll & Rounsaville, 2007; Dorsey et al., under review). Furthermore, in most community-based settings, the full gold standard "package," and particularly observation of practice, may not be feasible or sustainable. However, more practical strategies such as fidelity and symptom monitoring and/or behavior rehearsal may be sufficient to improve clinician fidelity and enhance client outcomes.

We propose a randomized trial that examines the effects of systematically including the above three gold standard supervision strategies in community-based settings. The "treatment platform" for this study is Trauma-focused Cognitive Behavioral Therapy (TF-CBT; Cohen, Mannarino, & Deblinger, 2006; Dorsey & Cohen, in press), an EBP for child trauma exposure and Posttraumatic Stress Disorder that has been widely disseminated (e.g., 18 statewide initiatives). The ongoing state-funded Washington State TF-CBT Initiative (involving PI Dorsey & Co-I Berliner) has specifically included provision of implementation training and support (i.e., consultation, listserv) for community-based supervisors. The proposed study builds on this initiative, offering a unique opportunity to study supervision of an EBP in community-based settings. We propose a two-phase, mixed within- and between-subjects design (Supervisors: *N*=20; Clinicians: *N*=140). In *Phase I: Baseline and Descriptive Study* (months 6-15), we examine current supervision strategies, particularly use of any gold standard techniques (i.e., supervision baseline), given the implementation support already provided to WA supervisors. In *Phase II: Randomized Trial* (months 15-39), clinicians are randomized to one of two supervision conditions, both of which systematically include additional gold standard strategies: a) Symptom and Fidelity Monitoring (SFM) or b) SFM plus Behavioral Rehearsal (SFM+BR). In SFM, the focus is on systematic monitoring of client symptoms and fidelity. In SFM+BR, the addition of BR provides a strategy that is potentially both a feasible proxy for observation of actual practice and an opportunity for skill building. This design allows for examining the impact on clinician fidelity and client outcomes of including gold standard elements, with the added benefit of a within-subject design, in which clinicians serve as their own controls prior to randomization. This innovative methodology investigates multiple salient factors in a single trial (e.g., implementation & client

outcomes) (Pringle et al., 2010), will generate findings relevant to a range of interventions, and will inform future TF-CBT implementation efforts.

Primary aims of this R01 are to: (Aim 1) Describe and establish baseline use of supervision strategies, when provided with implementation support (Phase I); (Aim 2): Evaluate effects of supervision condition (SFM vs. SFM+BR) on clinician fidelity, and in turn, client outcomes (Phase II); and (Aim 3): Test fidelity as a mediator for the relationship between supervision condition and client outcomes.

A secondary aim is to: Explore the relationship between supervision condition and broader implementation factors, including feasibility, acceptability, and impact on agency-relevant outcomes (e.g., turnover intention, burnout, supervision alliance).

Primary Hypotheses: (1a) Clinician fidelity will be higher for SFM+BR than for SFM (Phase II conditions), and higher for SFM and SFM+BR than for supervision in Phase I; (1b) Client outcomes will be better for SFM+BR than for SFM, and better for SFM than for supervision in Phase I; (2) The impact of supervision condition on client outcomes will be mediated by fideli

3. RESEARCH STRATEGY

3A. SIGNIFICANCE

A substantial number of evidence-based practices (EBPs) for child and adolescent mental health disorders exist, many with multiple RCTs supporting efficacy in improving outcomes and functioning (Weisz et al., 2005). However, outcomes for those served in community-based, "real-world" settings have been less positive (Weisz et al., 2006). Initial efforts to bridge the gap by providing access to EBPs through clinician training were mostly insufficient, as they did not include a focus on other key implementation factors, resulting in little to no change in clinician practice (Fixsen et al., 2005) and EBP implementation in community-based settings proceeding at "an unacceptably slow pace" (Mitchell, 2010, p. 2). Two recent reviews of clinician training that examined varying combinations of training elements indicate that training efforts must be active, multi-component, and attend to contextual factors such as organizational support (Beidas & Kendall, 2010; Herschell et al., 2010). One organizational support factor highlighted in both reviews as both important and receiving little research attention is supervision. According to Herschell et al. (2010), *the field requires "a better understanding of how supervisors should be trained and included in the implementation process"* (p. 463).

The role of community-based supervisors (CBS) in supporting EBP has been particularly overlooked, even though supervision is commonly provided in community-based settings, and therefore offers an existing mechanism for EBP support (Schoenwald et al., 2009). Very limited research, most using non-experimental designs, has focused on how CBS can feasibly and effectively support clinicians in providing EBPs (see A2). This is a critical gap in the implementation literature. Purveyors of EBPs have varying initial implementation requirements. Many require a period of model-specific supervision or consultation, typically provided by external experts (Forehand, Dorsey et al., 2010). With few exceptions (e.g., Multisystemic Therapy, Functional Family Therapy), external expert involvement is of limited duration. However, Beidas and Kendall (2010) conclude that "*ongoing supervision may be needed for actual therapist behavior change and skillful implementation*" (p. 3). For most community-based agencies, ongoing expert consultation is too costly to be sustained (Herschell et al., 2010; Kazak et al., 2010). In a unique EBP implementation study, inadequate funding was the primary reason for de-adoption (Massatti et al., 2008). For many agencies, more cost-effective solutions are needed (McHugh et al., 2009). In the limited research on sustainability, ongoing support, potentially in the form of supervision, is a critical factor (e.g., Blasinsky et al., 2006; Tibbets et al., 2010).

A1. "Gold Standard" Supervision Strategies

In efficacy trials and expert consultation models, a relatively common set of "gold

standard" supervision strategies has been used to support clinician fidelity (Milne, 2008; Padesky, 1996; Sheidow et al., 2008). Approaches typically include some combination of four elements (listed below), but currently there is little empirical guidance around which are critical, which combinations are most effective (Freeston, 2010), and which are feasible and acceptable in community-based settings.

1. Fidelity Monitoring. Fidelity, defined as "the degree to which an intervention was implemented as it was prescribed in the original protocol as it was intended by the program developers" (Proctor et al., 2010, p. 69), can be monitored via clinician self-report, observation of clinical encounters, and client-report. Monitoring fidelity efficiently and effectively is challenging (Hayes, 1998), creating tension between validity and cost (Schoenwald, 2011; Sheidow et al., 2008). The literature suggests that clinician self-report has poor concordance with objective ratings (e.g., Hurlburt et al., 2010). Objective measures (e.g., direct observation) although potentially more accurate, are also higher in cost and time (Hrisos, 2009). However, clinician-report, particularly when paired with other gold-standard elements, may be an effective strategy for enhancing the "fidelity-focus" of supervision (Amaya-Jackson et al., 2010; Henggeler et al., 1997). **2. Symptom Monitoring.** Regular monitoring of target symptoms, or "measurement-based care," has demonstrated potential for positive effects (Bickman, 2008), particularly when combined with fidelity monitoring (Chorpita et al., 2008; Seidman et al., 2010). In most efficacy and effectiveness trials, ongoing symptom monitoring (SM) is a necessary component of supervision, providing a means for regularly assessing client functioning and response to the intervention (Worthen et al., 2007). In the area of adult depression, a SM program implemented in over 500 community-based medical clinics nationwide was a highly effective tool for supervision (Hunkeler et al., 2006; Unutzer et al., 2002), as it allowed the supervisor (and clinician) to monitor client response even in the absence of direct client observation. **3. Skill Building/Behavioral Rehearsal.** Active, experiential learning strategies that build skills, typically through role plays with feedback, are an important supervision strategy for improving practice (Beidas & Kendall, 2010; Cross et al., 2007; Matthieu et al., 2007; Milne, 2009). Behavioral rehearsal (BR) also provides a proxy for observation of actual practice (see #4 below), in that the supervisor can observe demonstration of core treatment elements (Freeston, 2010). In medical education, BR with simulated patients *is considered* a form of direct observation (Hrisos et al., 2009; Peabody et al., 2000), yet has received less research attention as an observation proxy in mental health. BR may have less context validity than observation, but it is likely a feasible, multipurpose strategy for both assessing fidelity (via proxy) and improving skills (Matthieu et al., 2007). When combined with other strategies—self-report of fidelity, SM—BR may be particularly cost-effective and easily deployed in implementation. **4. Observation of Actual Practice.** Observation of actual practice (direct or via tape-review) is commonly used as a supervision strategy in efficacy trials, some effectiveness trials, and for some EBPs with ongoing requirements (e.g., Functional Family Therapy). Observation may be the most objective; however, many community-based agencies consider it too costly and resource-intensive to use consistently (Sheidow et al., 2008) and it is "*a method that differs considerably from most psychotherapy supervisi*on" (Schoenwald et al., 2009). In our work with CBS in WA (Dorsey, Berliner, Kerns), other states (Deblinger, Garland, Weisz) and nationwide (Unutzer), observation appears the least feasible and least likely to be systematically used, despite its benefits.

A2. The Potential Role of Community-based Supervisors in Supporting EBP Delivery

The majority of community-based mental health settings provide some form of ongoing clinical supervision (Schoenwald et al., 2008), yet the role of CBS in supporting EBPs has been largely overlooked. Local supervision, unlike external expert consultation, is generally covered within existing organizational and funding structures. Currently, very little is known about the focus and strategies used in CBS (Miller et al., 2006; Shoenwald et al., 2009). What is known suggests that supervision varies considerably and rarely, or inconsistently, includes "gold

standard" strategies (see C4 for preliminary studies by our team). Furthermore, because most clinicians are the direct service providers, most dissemination and implementation (DI) efforts predominantly or exclusively focus on clinicians despite typically high turnover. Supervisors, although providing less direct service, have significant influence on clinicians and may have lower rates of turnover. To facilitate DI, Chorpita and Regan (2009) suggest that we may need to *advance beyond examining practices with real-world clinicians and clients to examining "real-world supervisors and managers"* (p. 3). Supervision is potentially the "least investigated" aspect of implementation (Ellis et al., 1996; Kilminster & Jolly, 2000) despite findings that it accounts for a significant proportion of variance in client outcomes (Callahan et al., 2009). Supervision support may be even more important than training; in a recent study of training, the dose of model-specific supervision received (vs. training approach) predicted fidelity post-training (Beidas et al., under review). In a recent study of "usual care" in community settings, EBP-trained clinicians used EBP content and techniques at an insufficient dose to result in improved client outcomes (Garland et al., 2010). As summarized by Hershell et al. (2010), "*there does not seem to be a substitute for expert consultation, supervision, and feedback for improving skills and increasing adoption*." (p. 462).

A3. Supervision, Fidelity, and Client Outcomes

The research literature generally supports a link between *model fidelity and client outcomes* (e.g., Barber et al., 1996; Shoenwald et al., 2008; see Webb et al., 2010 for an exception) although the strength of the relationship varies. Stronger relationships between fidelity and outcomes are seen in effectiveness trials, likely because of ceiling effects in fidelity for efficacy trials (McHugh et al., 2009). In a recent effectiveness trial of Multisystemic Therapy (MST; 45 programs, 48 therapists, 1,979 families), fidelity was associated with better client outcomes (Shoenwald et al., 2008). The relationship *between supervision and client outcomes* is limited, but critical for supervision-focused studies, as improved client outcomes are the "acid test" for defining good supervision (Ellis & Ladany, 1997). In a CDC-funded RCT of SafeCare (MH065667; PI: Chaffin), only the monitored condition (1x monthly observed practice) achieved positive client outcomes (i.e., reduced rates of neglect). In non-experimental studies of MST using CBS, supervision strategies focused on adherence (i.e., as one aspect of fidelity) and on clinician skill development—but not those focusing on general support—were associated with client outcomes (Schoenwald et al., 2003, 2009). In one of the few studies that examined links between *specific supervision elements and fidelity*, only supervisor expertise in the EBP was associated with fidelity (Hennegler, 2002). Studies that experimentally tested gold standard supervision strategies in community settings were not identified. Learning appears to occur in supervision, as the "dose" of model-focused supervision may be associated with higher levels of fidelity (Beidas & Kendall, under review) and discontinuation of model-focused supervision may attenuate clinician skill/fidelity (Moyers et al., 2008). There is also evidence that *supervision is related to broader implementation outcomes* beyond fidelity. Specifically, supervision has been associated with decreased burnout or emotional exhaustion (Aarons et al., 2009; Knudsen et al., 2008), turnover intention (Knudsen et al., 2008), and actual turnover (Aarons et al., 2009).

A4. Proposed Study: Improving Practice through Focusing on Supervision

To leverage the potential positive impact of using CBS to support EBP, ***we plan to experimentally evaluate the impact of gold standard supervision strategies on clinician fidelity,*** *and, in turn, on client outcomes and broader implementation outcomes*. The study builds on a state-funded EBP implementation effort, the Washington State Trauma-focused Cognitive Behavioral Therapy (TF-CBT) Initiative (WA TF-CBT

Initiative), that has a specific, but limited, focus on training supervisors. The WA TF-CBT Initiative is an ongoing TF-CBT DI effort and EBP for child trauma exposure and Posttraumatic Stress Disorder (PTSD; Cohen, Mannarino, & Deblinger, 2006). TF-CBT has nine randomized trials supporting its efficacy (for reviews, see Dorsey et al., 2011; Dorsey & Cohen, [in press, App. D]) and has been widely disseminated. TF-CBT involves 12-20 sessions that include some child only, parent only, and conjoint sessions through which nine components are delivered (see Table 1). In efficacy trials, clinicians who received the gold standard TF-CBT training (2-day training, supervision involving regular, systematic use of 3-4 gold standard supervision strategies) resulted in positive client outcomes, including reduced PTSD, depressive symptoms and behavioral problems. Since the proposed study builds on an initiative that is similar to other state-level DI initiatives (see C1, Table 2), findings will generalize to other large-scale EBP DI efforts.

Table 1. TF-CBT Components (PRACTICE)	
Psychoeducation	**T**rauma Narrative & Processing
Parenting	**I**n Vivo Exposure
Relaxation	**C**onjoint Sessions
Affective Regulation	**E**nhancing Safety
Cognitive Coping	

Existing TF-CBT DI efforts often include a specific focus on CBS (Cohen & Mannarino, 2008). In NV and MS, supervisors were trained prior to clinicians. In CA and WA, supervisors attend both the basic and a supervisor-specific TF-CBT training. In states (NC, SC, & CT) that used the Institute for Health Improvement learning collaborative model (Kilo, 1998), a supervisor "track" provides specific training and discussion of supervision issues. However, the impact of training CBS on fidelity and client outcomes, for TF-CBT or other EBPs, has not been assessed. Currently, the field has "very limited data on supervisory practices" (Beidas & Kendall, 2010, p. 11) and little guidance on "what works." According to Garland et al.(2010), defining "what works" in implementation research may need to be interpreted not only as efficacy but also as "*practical, feasible, and affordable, and therefore, what is effective*" (p.16). To maximize the impact of efforts related to supervision training and strategies on clinician fidelity, the field needs practical recommendations taking efficacy, efficiency, and feasibility into account to specifically address questions such as: When supervisors receive some training and implementation support, which strategies are utilized? Which gold standard elements, of those most easily integrated into community settings, predict better fidelity and client outcomes?

The proposed study has significance for both implementation science and public health, in that the ultimate goal of improving supervision in community-based mental health settings is to increase access to EBPs for children and adolescents through improving clinician fidelity to TF-CBT. The potential public health impact of wide-scale provision of EBPs cannot be realized when EBPs are not sustained post-initial efforts. Increasingly, federal and state-funded EBP implementation efforts are underway, and in some cases, states and counties mandate EBP use (Bruns & Hoagwood, 2008; Jensen-Doss et al., 2009). Given the prevalence and high cost of these initiatives, *particularly if ineffective for changing and sustaining clinician practice,* it is critical to identify strategies for improving implementation and sustainability.

3B. INNOVATION

The proposed project is innovative at three levels. First, it *addresses a gap in the implementation literature* by focusing on how to improve supervision with the primary goal of enhancing clinician delivery of EBPs (i.e., TF-CBT) in community-based settings. Second, we propose to accomplish two goals in a single trial (Pringle et al., 2010), with a clear focus on *identifying practical, feasible recommendations that will have significant public health impact* (Strategic Objective 4; NIMH, 2008): 1) *describe existing supervision* (Phase I), given currently provided state-funded implementation support, and 2) *experimentally examine the effectiveness of systematically including gold standard supervision strategies* on clinician fidelity (Phase II). Strategies were specifically selected for both potential effectiveness ***and*** feasibility in community-based settings. This two-phased design will yield findings early in the trial about the focus and strategies of supervision with implementation support and, towards the end of the

trial, will yield information about the impact of supervision enhancements to support EBP (i.e., systematic use of gold standard strategies) on clinician fidelity and client outcomes. The proposed study is aligned with the goals of practical clinical trials (Tunis et al., 2003), as it will *provide information directly relevant to decision makers* (e.g., policy makers, payers). This study will be the first to specifically test the addition of behavioral rehearsal, a multipurpose strategy, in community-based settings. Third, the proposed study *leverages novel technological approaches* (see C5a) (e.g., i-Touch devices) to implement study procedures (e.g., audiorecording, data collection) and to support the Phase II supervision conditions (e.g., house SFM & BR aspects). According to Pringle et al. (2010), "implementation studies too often neglect the potential contributions of emerging technologies to help integrate effective practices within care systems" (p.194-195). This study uses a relatively low-cost mobile platform (i.e., handheld devices) that holds promise for greater use of technology in the diverse settings in which mental health care is increasingly provided (e.g., in-home, school, residential settings). Given innovation at multiple levels, we expect that the proposed study will considerably advance the field in the area of implementation of EBP and offer viable options for improving care for children and adolescents.

Table 2. TF-CBT State Initiatives (5 of 18)	ND	WA	CA	SC	DE
Est. cost/yr (in thousands)	125	100	1000	800	400
Web training	✓		✓	✓	✓
2-day basic training	✓	✓	✓	✓	✓
6- mo expert consultation	✓	✓	✓	✓	✓
Adv TF-CBT training	✓	✓	✓	✓	
Sup training/consult		✓	✓	✓	
Admin training		✓		✓	

3C. APPROACH

The ultimate goal of this study is *to provide practical recommendations about supervision strategies that support clinician fidelity in community settings*. As detailed in Significance, research on supervision is limited, with little data to guide recommendations about which gold standard elements can (i.e., feasibility) and should (i.e., effectiveness) be used.

C1. Washington State TF-CBT Initiative (WA TF-CBT Initiative)

The structure for the WA TF-CBT Initiative is similar to the other 17 statewide Initiatives (Cohen & Mannarino, 2008; Sigel et al., 2010; see Table 2) and to other national or statewide DI efforts (e.g., NY's Achieving the Promise for Children, Youth, and Families; Bruns & Hoagwood, 2008). Participation in the state-funded WA Initiative has been geographically diverse, with involvement from over 60 agencies statewide. Since 2007, nearly 700 clinicians and supervisors have received training in TF-CBT. Participation involves an agency administrator attending a half-day training; clinician-supervisor teams receive the manual (Cohen et al., 2006), attend a 2-day in-person didactic and experiential training; and receive 6-mo of biweekly, 1-hour, expert consultation by telephone. Clinicians can also attend a 1-day advanced training.

C2. WA TF-CBT Initiative Participants: Representative of Other Community-based Providers

As demonstrated in a study conducted by our team (Dorsey, Lyon, Pullmann, & Berliner, under review), WA TF-CBT Initiative participants are highly representative of community-based providers in other studies focused on real-world settings (e.g., Garland et al., 2010). Supervisors and clinicians are predominantly female, have master's degrees, and identify an eclectic orientation (see Table 3 & Figure 1 for descriptive data on most recent cohorts: 2009 & 2010).

Table 3. WA TF-CBT Initiative Descriptive Data: Two Most Recent Cohorts (2009 & 2010)		
	2009 (*N* = 72)	2010 (*N* = 92)
Demographics		
Female	61 (84.7%)	75 (83.3%)
Age		
25-29	13 (18.3%)	21 (23.3%)
30-39	31 (43.7%)	41 (45.6%)
40 and older	26 (36.4%)	25 (27.7%)
Education		
Masters	63 (86.1%)	84 (93.3%)
PhD	2 (2.8%)	2 (2.2%)
Exper. (Yrs)	*M*=7.16 (SD=6.9)	*M*=5.86 (*SD*=6.1)

C3. Initiative Training: Improved Report of TF-CBT Ability

Findings from the 2009 cohort demonstrate significantly improved *perceived ability* in assessing trauma and PTSD and in delivering TF-CBT components, achieving a mean overall TF-CBT competency score of 73% (Dorsey et al., under review), which falls between the 70%–80% "proficiency standard" scores used in recent clinician training reviews (Beidas & Kendall, 2010; Beidas et al., under review). While these scores are in the acceptable range, there is still room for improvement. As scores are *self-report of ability* rather than objective ratings, they are likely overestimates (Beidas & Kendall, 2010). Also, fidelity attenuates over time without ongoing supervision (Moyers et al., 2008). Hence, although WA TF-CBT Initiative training has been sufficient for achieving reasonable perceived ability, effective supervision is necessary to build and sustain EBP skills. However*, research has yet to clearly identify effective and feasible supervision strategies for CBS.*

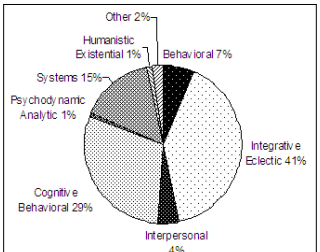

Figure 1. Theoretical orientation of clinicians participating in WA Initiative.

C4. Preliminary Studies Show Inconsistent Use of Gold Standard Elements

To provide additional implementation support and encourage CBS to support TF-CBT, supervisors in the WA TF-CBT Initiative are invited to attend a 1-day supervisor training. This training (led by Dorsey) reviews supervision challenges for TF-CBT and EBPs and introduces gold standard supervision elements (e.g., fidelity/symptom monitoring). Supervisors also are invited to participate in a 1x monthly (1 hr) supervisor call (led by Berliner) and a supervisor-only listserv. *However, this supplemented "train and hope" approach to supporting supervisors appears to be insufficient.* In an assessment of supervision practices at the 2010 supervisor training (*N=22),* supervisors reported that they *provide EBP-focused supervision* (*M* = 3.4 "most" on a 1-4 scale, see Table 4); but supervisors reported *only occasional use of fidelity monitoring* (*M* = 2.3), with only one supervisor reporting that fidelity monitoring was a routine activity (Dorsey et al., under review).

Table 4. WA TF-CBT Initiative Supervision Strategies

	Never	Occasionally	Most of the Time	Ongoing/ Routine
EBP supervision	-	3(13.6%)	7(31.8%)	12(54.5%)
Fidelity monitoring	2(9.1%)	11(50.0%)	8(36.4%)	1(4.5%)

In a second preliminary study involving CBS by our team (Accurso, Taylor, & Garland, 2011) (supervisors: *N* = 7; clinicians: *N* = 12), supervisors commonly reported using progress notes review in supervision (83.2%) yet reported infrequent use of gold standard strategies (i.e., 4.6% fidelity monitoring; 13% tape observation). According to supervisor and clinician report, supervision occasionally included some discussion of EBP-related content and techniques; however common EBP techniques, like assigning/reviewing homework, were discussed *in less than half* of the meetings. These studies clearly indicate that <u>although CBS may occasionally use gold standard strategies, particularly when some implementation support is provided, *these strategies do not appear to be used consistently or in sufficient "dose" to change clinician practice*</u>.

C5. Overview of Study Design, Sample, and Data Collection

Our preliminary studies provide important information about CBS, but studies are in limited in scope and rely on self-report (vs. objective measurement). We propose a two-phased approach to: a) first *identify strategies supervisors already use*, given existing levels of implementation support (e.g., 1-day training, monthly call, listserv); and b) *study the impact of systematically and consistently including gold standard elements* in supervision.

<u>Experimental Design.</u> Phase I is descriptive and involves establishing a baseline for current practice, specifically identifying use of any gold standard strategies; Phase II involves training supervisors and randomizing clinicians to one of two supervision conditions.

During Phase I (months 6-15, see Table 5), we examine current supervision practice, hereafter referred to as "Supervision with Implementation Support" (SIS), given the supervision training and support currently provided as part of the WA TF-CBT Initiative. All supervision

meetings will be audiorecorded and a random sample will be coded for content (e.g., TF-CBT components, administrative topics), techniques (e.g., behavioral rehearsal, observation of actual sessions, use of fidelity monitoring tools), and intensity/depth of content and techniques (see C8 for procedures & Supervision Coding Measure details, see also App. B). Studying SIS, current supervision practice, is critical. Implementation support provided to supervisors as part of the WA TF-CBT Initiative is equal to, or exceeds, most statewide initiatives for TF-CBT (see Table 2; Sigel et al., 2010) and other EBPs. The Phase I study addresses a gap in the literature by describing CBS, particularly any use of gold standard elements, provides a baseline for supervision, and improves upon our preliminary studies of CBS practice by increasing scientific rigor (i.e., objective measurement).

During Phase II (months 15-49), all clinicians will be randomized to one of two supervision conditions: Symptom and Fidelity Monitoring (SFM) or SFM plus Behavioral Rehearsal (SFM+BR) (see C6 & C7 for detailed descriptions of conditions and supervisor training). We randomize by clinicians who are also the unit of analysis, given our ultimate goal of improving clinician fidelity. Supervisors will provide both conditions to reduce impact of individual supervisor/agency factors (see Section C11 for consideration of other designs & C8 for protections against experimental drift). Phase II conditions allow for controlled study of supplementing SIS with: 1) *systematic monitoring of client symptoms and TF-CBT fidelity* and 2) *behavioral rehearsal* on clinician fidelity, client outcomes, and broader implementation outcomes (e.g., burnout; turnover intention). Fidelity scores will be obtained from 3 clients per clinician to obtain precise and reliable estimates of effect (see C10).

C5a. Leveraging Technology for Study Procedures and Data Collection

Advances in technology offer opportunities in improving the ease and availability of real or near real-time, technical support for treatment delivery and research (Pringle et al., 2010). We leverage these advances by utilizing low cost mobile devices (e.g., Apple i-Touch) to audiorecord supervision and client sessions (devices have native audio recording programs, 20-40 hours capability), to house the SFM and BR materials, and to facilitate completion and timely receipt of clinician and supervisor weekly-completed client session and supervision meeting forms. The majority of participating agencies (85%) have wireless internet, allowing regular and secure uploading of recordings, monitoring data, and weekly session/supervision data (see Human Subjects for detailed information on mobile device security). The decision to provide portable devices over accessing web-based program via a laptop/desktop computer is based on our experiences with clinicians in the WA TF-CBT Initiative, one-third of whom did not have access to a computer *during* client sessions (prohibiting a system that requires computer access). Clinicians have *daily* computer access, allowing for wired device syncing and uploading for those at agencies without wireless capability (15%). Devices also offer a benefit over paper-based data collection in that they allow for graphing symptoms for review with supervisors and centralized electronic data storage and uploading, resulting in reduced research team tracking. CY Consulting (CYC; see letters of support), a technology and strategy consulting group with specific experience in healthcare settings, will work with our team during the Startup and Phase I to develop applications, as devices will only be used to audio record sessions, provide session data, and access online surveys until Phase II starts. Development of the SFM and BR applications will involve biweekly phone/Skype meetings between CYC, the PI, and the Project Coordinator. Dr. Unutzer will also participate in these meetings on a 1x monthly basis, as his expertise in designing the monitoring system used in the IMPACT trials will inform development. CYC recently developed, and successfully launched, a range of Apple mobile applications (i-Touch, i-Pad) for providers in a large children's hospital in CA—allowing our team to build on existing healthcare application development experience. To ensure high levels of clinician and supervisor satisfaction and usability, supervisors and clinicians from an agency in CA (see letters of support) will pilot the applications and provide feedback for 2 iterative development cycles during the development period.

Table 5. Study Activities Timeline						
ACTIVITIES	**YEAR 1**		**YEAR 2**	**YEAR 3**	**YEAR 4**	**YEAR 5**
Study Phase :	**Startup**	**Phase I**	**Phase II**			
i-Touch application development						
Refine supervision coding						
Pilot supervision coding measure						
Clinician enrollment						
Supervisor/clinician assessment						
Study training						
Youth enrollment						
Youth assessment						
Supervision coding						
TF-CBT fidelity coding						
Clinician randomization						
Supervision training						
Data analyses; Manuscript prep.						

C6. Description of Phase II Supervision Conditions

Symptom and Fidelity Monitoring (SFM). Symptom Monitoring involves clinicians monitor key symptoms each session using brief measures via the i-Touch (anxiety/PTS; externalizing behavior; see Table 7 for a list of measures; see App. B for actual measures). Using the program created for this study, which builds on Dr. Unutzer's monitoring system used successfully in the IMPACT trials (Unutzer et al., 2002), symptoms will be scored and graphed over time/sessions to provide a pictorial representation of improvement or deterioration, and reviewed in supervision. For the *Fidelity Monitoring* clinicians complete a brief, standard TF-CBT checklist after each session (via i-Touch), which is also reviewed in supervision. The supervisor will have a slightly elaborated version that includes follow-up queries for TF-CBT components and key CBT techniques (assigning/reviewing homework, role plays) and a cue to plan for the upcoming session (see App. B). The goal is to determine if systematic monitoring and review—of symptoms and fidelity—is sufficient to improve fidelity and client outcomes. Potentially, even when not observing the clinician's practice (through observation or role play), closely monitoring fidelity and client response (i.e., symptoms) may be sufficient for improving clinician fidelity. If so, SFM presents a feasible, scalable supervision-improvement intervention that could be employed in implementation efforts for a range of EBPs.

Monitoring and Behavioral Rehearsal (SFM+BR). The SFM+BR condition includes the SFM components and an additional, multipurpose strategy—*behavioral rehearsal (BR)*. Given concerns about relying on clinician self-report of fidelity and the importance of skill building, the inclusion of BR serves two purposes, making it a potentially high-yield addition. First, it provides a proxy for actual practice (e.g., observing skill via role play). Second, it allows the supervisor to address ongoing skill building by having the clinician practice techniques and receive supervisor feedback. In lieu of observing actual practice, BR may offer considerable benefits and, compared to observation of actual practice, is more likely to be easily incorporated in community-based settings. In SFM+BR, supervisors also will conduct a TF-CBT component BR relevant to an upcoming session (e.g., 5-10 min) and provide feedback. Supervisors will have a short set of guidelines (e.g., bulleted form) that detail expected content and techniques for BR of each TF-CBT component (see App. C). Guidelines focus specifically on common challenges for TF-CBT components (e.g., dealing with avoidance, homework assignment/review).

Phase II supervision conditions are not "one size fits all." Both focus on fidelity to TF-CBT, but also on individual client response. TF-CBT is inherently flexible. Fidelity involves providing all 9 components (see Table 1), but delivery can, and should be, adjusted to fit the child's particular developmental and/or cultural needs. Interestingly, the SFM+BR condition may hold the most promise for improving cultural fit, in that supervisors can better coach and plan with clinicians to provide TF-CBT in a culturally-sensitive, high-fidelity manner.

C7. Training with Supervisors and Clinicians

All participating supervisors have been involved with the WA State TF-CBT Initiative for 1-5 years, have been providing TF-CBT to clients, and have demonstrated particular TF-CBT leadership at their agencies (see Table 6, next page). Supervisors receive additional training in Phase II of the study, on the experimental supervision conditions, but all participating supervisors have been receiving SIS for 1-5 years.

Phase I: Supervisors and Clinicians. At the beginning of Phase I, clinicians and supervisors will participate in a 2-day study training focused on youth recruitment and mobile device training (e.g., session audiorecording & data collection [i.e., self-report measures; weekly session information]).

Table 6. WA State TF-CBT Initiative Supervisor Characteristics

Characteristics	Mean/ %
Years Experience	14.3
Years at Agency	10.5
Years in WA TF-CBT Initiative	3.9
Number Supervisor Trainings Attended	1.85
% Regular/Frequent Participation in WA TF-CBT Supervisor Calls	69%
% with ≥ 10 TF-CBT Cases	61.5%

Phase II: Supervisors Only. In the beginning of Phase II, supervisors (no clinicians) will participate in a 2-day training on the Phase II supervision conditions and study procedures (See App. E, Supervisor Training Agenda). Drs. Dorsey, Deblinger, and Ms. Berliner will provide training in how to systematically integrate gold standard elements into supervision for the Phase II conditions. Training will include didactic information, benefits of SFM and BR, modeling of SFM and SFM + BR use, and coached supervisor practice with feedback on SFM and SFM+BR procedures and use in supervision. Trainers will provide clinician vignettes and pre-loaded SFM data on the i-Touch to practice implementing the experimental condition and using the application. For SFM+BR practice, vignettes also include a guide for the supervisor (who plays the clinician) to make the role more realistic. After practice, to reinforce future use, supervisors will be asked to discuss how they think use of the strategies can enhance their supervision of TF-CBT. The training will also include research training on our experimental design and trial procedures with particular emphasis on the importance of condition adherence and preventing experimental drift (see C8: study drift protection). Trainers will clearly and repeatedly emphasize that experimental drift would severely compromise the study, such that study aims would not be achieved. Supervisors will practice identifying drift in vignettes of supervision meetings.

C8. Coding Procedures

Our team has significant experience with using audiocoding procedures as an objective measure for assessing fidelity to TF-CBT (Deblinger & Dorsey) and for capturing broader content and techniques, as well as "extensiveness" (i.e., time; thoroughness) (Garland; see Garland et al., 2010; PRAC Therapy Process Observational Coding System for Child Psychotherapy; TPOCS-S).

TF-CBT Fidelity Coding Procedures. All client sessions will be audiorecorded. Three sessions will be coded for each child, randomly selected from throughout treatment, using the TF-CBT Scoring Checklist (Deblinger et al., 2005). Coding involves capturing prescribed TF-CBT components (see Table 1; e.g., Psychoeducation; Relaxation; Trauma Narrative) in child only, parent only, and conjoint parent-child sessions. The coding includes well specified items, an existing coding manual (Cooper, Young, & Deblinger, 2008; see Table 7), acceptable levels of interrater reliability (.79 and .80 for child only, and parent only sessions, respectively; see App. B for a copy of the measure), and has been used successfully in previous TF-CBT RCTs. *For all coding (TF-CBT and Supervision), research assistants will be blind to study condition.*

Supervision Coding Procedures. All supervision meetings with study clinicians will be audiorecorded using the mobile devices. In Phase I, 20 meetings per supervisor, randomly selected from Phase I, with selection distributed across participating clinicians will be coded (total of 400 meetings). During Phase II, four randomly selected meetings per supervisor (2 for each condition, distributed over time across participating clinicians), will be coded each month within a two-week window from when the supervision meeting occurred (possible due to near

real-time uploading with use of the mobile device system). Research assistants will use the TF-CBT Supervision Coding Measure (see below for measure description, Table 5, & App. B) developed for this study, which will be refined with assistance from Dr. Garland (consultant) during the startup period.

TF-CBT Supervision Coding Measure. Given limited supervision research, no existing coding measure was identified that, like the TPOCS-S for child therapy, captures *both:* 1) general supervision content and strategies and 2) EBP (e.g., TF-CBT) content and strategies. The content and technique items on the supervision coding measure for the proposed study were identified by review and synthesis from: 1) an existing, psychometrically, valid CBT-focused coding measure (i.e., Supervision: Adherence & Guidance Evaluation; Milne & Reiser, 2008); 2) a recent review of supervision interventions (Milne et al., 2008); and 3) a self-report supervision questionnaire used in Dr. Garland's CBS pilot study (Accurso et al., 2011). Similar to the TPOCS-S, the resulting coding measure includes coding of both content of supervision (e.g., TF-CBT components; administrative paperwork requirements; client homework) and strategies used in supervision (e.g., symptom review; case note review; behavioral rehearsal), as well as extensiveness (e.g., time spent? thoroughness?). During the startup period, the supervision coding measure will be refined (i.e., any additional strategies/content identified & added) in collaboration with Dr. Garland, then piloted with 2 supervisors in a CA community-based mental health clinic (see letters of support). The coding measure will allow for capturing gold standard strategies used in SIS (Phase I) and assessing adherence to condition in Phase II.

Training Procedures. Research assistants (RAs; see Budget Justification; Personnel Section) will be trained in coding procedures for client sessions and supervision meetings during the startup period. Training will be conducted by the PI and Project Coordinator, with assistance from Drs. Deblinger and Garland. Training will include didactics, manual review, practice sessions, and training to an established criterion (80% agreement on training tapes). During the study, a random 20% of sessions will be coded by two RAs to test interrater reliability. If reliability is < 80% on 2 successive reviews, booster training will be conducted.

Experimental Drift Protections. On a weekly basis, coding results will be reviewed by Drs. Kerns or Pullmann to check for supervisor adherence to condition. If experimental drift is identified (e.g., supervisor incorporates BR into the SFM condition; no use of either/both monitoring strategies in either condition), supervisors will be contacted within two weeks for a booster training on the supervision conditions via study-purchased web camera. Booster training will include didactic review, role plays of SFM and SFM+BR and a review of drift implications. In addition, supervisors will receive a biweekly email on condition adherence, based on Dr. Pullmann and Kerns' review of the data, so that all supervisors receive regular communication from our team related to provision of supervision conditions. Our goal will be to provide positive reinforcement for adherence, reminders about maintaining experimental integrity, and schedule booster trainings when needed.

C9. Participants: Recruitment, Retention, and Data Collection Procedures

C9a.Supervisors and Clinicians

Recruitment. Participants will include 20 supervisors and 140 clinicians from 15 agencies (see letters of support). Agencies were selected from those in the WA TF-CBT Initiative with consideration of statewide geographic representation. In an attempt to reach a more diverse provider and client population, we will specifically encourage participation from agencies in urban areas and in Eastern Washington, both are areas in WA with more greater ethnic and cultural diversity (however, we still conservatively estimate enrollment, as participation is voluntary; see Expected Enrollment Table). In the PI's ongoing study of TF-CBT with foster youth, conducted in *only* an urban area, the client population was highly diverse (i.e., $N = 49$; 53% multiracial, 25% Caucasian, 18% African American, 2% Native American, 2% Asian). Each agency has 1-2 supervisors and approximately 10 clinicians that meet study inclusion and

exclusion criteria. The number of clinicians supervised by each supervisor will vary, with Phase II conditions balanced across supervisors. Although number of clinicians will vary, supervision time is similar across agencies; clinicians receive 45-60 minutes of individual supervision per week. Enrollment will occur during the 6-mo startup period (see Table 5 & Human Subjects for more details). Some eligible supervisors have already been identified (but not enrolled); eligible clinicians must (a) be trained in TF-CBT as part of the WA TF-CBT Initiative and completed Initiative requirements (i.e., 2-day basic training; 6-mo expert consultation); (b) be at least 80% FTE employees at one of the 15 agencies; (c) currently provide TF-CBT to children and adolescents; (d) supervised by one of the participating supervisors; and (e) provide treatment in English. We propose few exclusion criteria, only (a) immediate plans to leave the agency or transition into a non-child/adolescent caseload carrying position.

Data Collection Procedures. Assessment measures for supervisors and clinicians will be completed prior to the study training in Phase I (A1; "baseline" [BL]), at the end of Phase I (A2: 9-mo post-BL), 1 year into Phase II (A3: 21-mo post BL), and at the end of Phase II (A4: 39-mo post-BL). Assessment of supervision practice, via coding supervision meetings, will be ongoing as described above (see C8).

Retention. To participate in *all* data collection points for Phase I and II, supervisors and clinicians would need to be retained for 3.4 years. For this study, supervisor attrition presents the most significant threat to achieving study aims. We expect very little attrition among *supervisors*, as WA TF-CBT Initiative supervisors have high levels of retention (see Table 6, *M* = 10.5 years). To prepare for potential supervisor attrition, however, we have invited a second supervisor (in the agencies in which there is not already a second supervisor) to participate in the 2011 WA TF-CBT Initiative training (March, 2011) and in the supervisor training (September, 2011) so that by study initiation (April, 2012) more agencies will have a second supervisor eligible for participation. As trainings are ongoing, we are also able to invite new agencies, supervisors, and clinicians to participate as needed. Clinician attrition will occur, and we have structured our design and analyses (e.g., larger than needed *N*, within & between subjects analyses, multiple assessments) to account for attrition. We expect low attrition in Phase I (e.g., 10%) due to its short duration (i.e., 9-mo). If attrition is greater than 15%, additional clinicians who meet eligibility criteria will be invited to participate at the beginning of Phase II. In this way, attrition in Phase I slightly impacts within-subject analyses (already more highly powered), but would not impact between-subject analyses in Phase II. Because we have a sufficient pool of clinicians and supervisors, as we are capitalizing on ongoing WA TF-CBT Initiative trainings, clinician turnover will not exceedingly impact Phase II between-subjects analyses. Phase II is 2.5 years. If we experience more than 50% turnover in Phase I, we will increase Phase II enrollment numbers to account for higher than expected turnover. We can also use clinician data that is less complete than expected (i.e., fewer numbers of clients) or incomplete (missing data on individual clients). Currently, we are powered for 20% attrition in Phase II, over and above 10% attrition in Phase I (70% retention overall; ultimate retained *N* = 50 in each Phase II condition). Clinician retention estimates are based on Dr. Dorsey's ongoing TF-CBT study that includes four agencies in the WA TF-CBT Initiative (clinician *N* = 18; 89% retention over 2.5 years) and on WA TF-CBT Initiative retention data for most recently completed cohort (*N* = 72; 84% retention). We conservatively lowered our retention expectations from rates found in our prior studies, given the current economic climate and higher data demand burden placed on clinicians in this study.

C9b. Youth and Parents

Recruitment. Eligible children and adolescents will be clients who present at one of the 15 agencies for mental health treatment. Inclusion criteria include: a) age 8-15; b) trauma history; c) significant PTS symptoms (i.e., score > 21 on the UCLA PTSD-RI (Steinberg et al., 2005); see Table 7), d) live with a parent (defined as a biological, adoptive, or foster parent) who is willing to participate in the study; and e) the child and parent speak English. Exclusionary

criteria are only a) presence of a pervasive developmental disorder or cognitive impairment (i.e., IQ $\leq$ 70) and b) parental serious mental illness (e.g., schizophrenia). Participating agencies have internal procedures for referring child clients with a trauma history and potential need for TF-CBT to clinicians trained in TF-CBT. As part of the WA TF-CBT Initiative, clinicians use the Child Posttraumatic Stress Scale (CPSS; Foa et al., 2001) to assess trauma impact. Clinicians will introduce the study to the caregiver of potentially eligible youth in the session in which the CPSS is administered and TF-CBT was deemed an appropriate treatment option (CPSS will not be shared with study staff). If permission is obtained, the clinician will provide contact information to the study team, who will then contact the family for eligibility screening and enrollment (see Human Subjects for details). The study team will notify the clinician and supervisor about enrolled cases to trigger initiation of relevant study procedures.

Data Collection Procedures. Data collection from youth and 1 parent is minimal. Measures will be completed at the beginning of treatment (BL), 3 mo into treatment (3M), and at the end of treatment (ET). For clients who drop out of treatment (passive = 4 weeks of no contact), the "end of treatment" assessment (to control for time) will be conducted at 6 mo. Only 2 measures will be administered (see Table 7; UCLA PTSD-RI; Strengths and Difficulties Questionnaire [Goodman, 1997]). Due to statewide distribution of agencies, measures will be completed by telephone, following procedures used in the Clinic Treatment Project Phase II and III trials (Weisz, PI: consultant on this proposal) for these same measures.

Retention. In Dr. Dorsey's ongoing TF-CBT study, retention rates for high-risk and mobile participants (i.e., foster care) are 89.3% (for the 28 youth, of 50, at end of treatment). In the 4 largest TF-CBT trials (i.e., Cohen et al., 2004; Cohen et al., 2005; Deblinger et al., 2010; Deblinger et al., 2006), average study retention rates for end of treatment data are 78.9% (range: 73.2% – 89.1%). Our research team has developed effective procedures for maintaining contact with participants (e.g., birthday/holiday cards to track change of address, multiple means of contact), even when treatment discontinuation occurs (active or passive). We expect clinicians to enroll one client in Phase I and three clients in Phase II (Total *N* = 560 youth) (see Expected Enrollment; Human Subjects for *N* feasibility). Given drop out rates in community mental health, we expect 65% retained *in treatment* (*N* = 364) through the end of treatment assessment. We expect higher rates of retention *in the study* (75%; *N* = 420).

Table 7. Study Measures (All Measures are included in Appendix B)

Domain; Respondent		Measures & Indicators	Interval
Client Outcomes	Posttraumatic Stress; Y, P#	*UCLA Post Traumatic Stress Disorder Reaction Index* (UCLA PTSD-RI; Steinberg et al., 2005). The 38-item UCLA PTSD-RI assesses trauma exposure and posttraumatic stress symptoms. The UCLA PTSD-RI demonstrates good convergent validity and good to excellent test-retest reliability, with Cronbach's α in the range of .90.	BL, 3M, ET*
	Overall Functioning; P, Y (over age 11)	*Strengths & Difficulties Questionnaire* (SDQ; Goodman, 1997). The SDQ is a short behavioral screening (ages 3-16) with 25 items on 5 scales (Emotional Symptoms, Conduct Problems, Hyperactivity/Inattention, Peer Relationship Problems, & Prosocial Behavior). The SDQ has been shown to have good discriminant validity, acceptable levels of test-retest reliability, and a Cronbach's α of .73 (Goodman, 2001).	BL, 3M, ET
	Demographics; P	*Family Information Form* (Child STEPS; Weisz & Chorpita, 2005). This form collects information on family, parent, and child demographics. John Weisz is a consultant on this project.	BL, 3M, ET
	Therapy Session Details; C	*Session Tracking Form* (Dorsey et al., under review). Developed as part of Dr. Dorsey's ongoing study of TF-CBT, this form collects session date, attendees, duration and information on any cancelled or rescheduled sessions.	Every session
Syx. Mon.	Anxiety & PTS; Y	*Screen for Child Anxiety-Related Emotional Disorders—Anxiety and PTS Subscales* (SCARED; Birmaher, et al., 1997; Muris et al., 2000). SCARED subscales are used for symptom monitoring in the Phase II Conditions. The SCARED is a 41-item measure; only general anxiety (5-items) and PTS (4-items) subscales are used in Phase II. †	Every session
	Externalizing Behavior Problems; P	*Pediatric Symptom Checklist-17; Externalizing Subscale only* (PSC-17; E) (Gardner et al., 2007). The PSC-17 is a 17 item measure that includes subscales for internalizing, externalizing, and attentional problems). The 7-item externalizing subscale is used for symptom monitoring in the Phase II conditions. †	Every session

Fidel.	TF-CBT Fidelity; RC	*TF-CBT Checklist Scoring Sheet* (Deblinger, et al., 2005). The TF-CBT Checklist Scoring Sheet was used in our prior TF-CBT studies and has a well-developed coding manual (Cooper, Young, & Deblinger., 2008) that will be used in the current study. Interrater reliability for the parent and child sessions were .80 and .79, respectively.	3 random selected sessions/youth
Fidelity Mon.	Self-report of Fidelity C	*Brief Practice Checklist* (BPC; Deblinger et al., 2008). The BPC is a clinician self-report checklist of TF-CBT components frequently in TF-CBT implementation efforts. Used in Phase II as a fidelity monitoring strategy.	Every session (Ph. II)
	Supervisor-report of Fidelity; S	*Brief Practice Checklist—Supervisor Version BPS-S).* The BPC-S, developed by our team, includes the BPC items and additional items for supervisors, including use of behavioral strategies uncommon in usual practice (modeling, role play, assign/review homework; Garland et al., 2010). The BPC-S was piloted by expert consultants in Dr. Dorsey's ongoing TF-CBT study and by two CBS in CA. Used in Phase II Conditions as fidelity monitoring strategy.	Every sup. mtg. (Ph. II)
Supervision	Supervision content, strategies, and intensity; RC	*TF-CBT Supervision Coding Measur*e. This measure will be used in Phase I to describe content and strategies for SIS and in Phase II to conduct "checks" on experimental condition. The measure includes general and EBP (TF-CBT) content and strategies. Dr. Garland will work with the research team in Phase I to finalize development of this measure and the coding manual (Garland et al., 2008) and to conduct a rapid pilot and revision with 2 TF-CBT supervisors in a community-based setting in CA (see EMQ-Families First letter of support).	Random selected meetings
	Supervisory Relationship, C	S*upervision Alliance Scale* (SAS; Knudsen et al., 2008). The SAS has 12 items; 5 drawn from Efstation et al. (1990); 7 from Rahim (1988). The 12-item measure demonstrated predictive value for burnout and turnover intention and had a Cronbach's α of .95 (Knudsen et al., 2008).	A1, A2, A3, A4^
	Feasibility/ Acceptability; S	*Supervisor Interview—Revised* (Kerns, Dorsey, Trupin, & Berliner, 2010). The SI is a qualitative interview and was used previously by our team to evaluate feasibility and acceptability of an expert consultation model for child welfare workers. The SI has been revised for this study to include questions specific to Ph II SFM and SFM+BR.	A4
Sup./Clinician Characteristics	Attitude towards EBPs; S, C	*Evidence-Based Practice Attitude Scale* (EBPAS; Aarons, 2004). The EBPAS has 15-items with 4 subscales: appeal, requirements, openness, and divergence. Cronbach's α for the EBPAS ranges from .66 (i.e., divergence) to .91 (i.e., requirements). The Cronbach's α for the total score =.76. The most recent EBPAS study shows that clinician scores are typically independent and provides norms reflective of a national sample (Aarons et al, 2010).	A1. A2, A3, A4
	Background Information; S, C	*Therapist Background Questionnaire* (TBQ; Weisz & Chorpita, 2005). Measure assesses prior training, education, theoretical orientation, and other background information. We added 4 items on participation in the WA Initiative and TF-CBT-specific items (e.g., year trained, TF-CBT web) and questions about training in other EBPs.	A1
	Organizational Climate; S, C	*TCU-Organizational Readiness for Change* (TCU-ORC; Lehman et al., 2002).The TCU ORC consists of 115 items on a 5-point scale. There are 4 main scales (motivation for change, adequacy of resources, staff attributes, and organizational climate) and 18 subscales. Cronbach's α for the 18 subscales range from .49 (change) to .84 (immediate training needs) with 10 of the 18 subscales having reliabilities above .70. (Lehman et al., 2002).	A1, A2, A3, A4
	Burnout; S, C	*Maslach Burnout Inventory, Emotional Exhaustion Subscale* (MBI-EE; Maslach et al., 1996) & item #8 on *TBQ* (see above). The 9-item MBI- EE, one of the most important factors for burnout and was used in the Knudsen et al. (2008) supervision study. Cronbach's α for the MBI-EE is .92 (Kim & Ji, 2009)	A1, A2, A3, A4
	Turnover; S, C	*Turnover Intention* (Knudsen, 2008). Assessed by 4 items adapted from Walsh et al. (1985) and used in Knudsen et al. (2008). Cronbach's α was .85. Actual turnover assessed by supervisor report at end of Phase II.	A1, A2, A3, A4

Gray Sections: Measures used for Phase II symptom and fidelity monitoring (SFM) and not used as study outcome measures.

#Y = Youth; P = Parent; S = Supervisor; C = Clinician; RC = Research Assistant Coder; *BL = Baseline; 3M = 3 months; ET = End of Treatment; ^A1 = Baseline (beginning of Phase I); A2 = End of Phase I; A3 = 1-year into Phase II; A4 = End of Phase II; †These measures are already used as a part of the Washington State TF-CBT Initiative.

C10. Data Analysis Plan

Basic data screening procedures will be conducted to screen for errors and explore normality, linearity, form, and outliers. Data will be transformed as appropriate. We will confirm randomization validity (e.g., chi-square tests, t-tests, Kruskal-Wallis). Minor differences will be statistically controlled during model-building. We will explore for selection bias from attrition. Data missing at random will be modeled using full maximum likelihood estimation. Quantitative analyses will be conducted using the SPSS (IBM-SPSS, 2007) and HLM (Raudenbush et al., 2009) software programs. We will assess "supervisor adherence" to establish that clinicians received supervision appropriate to assignment. Within-groups ANOVAs will explore changes in Supervision Coding Measure scores from Phase I to II, including frequency and length of supervision meetings, techniques used, and frequency/intensity of techniques. We expect little

change in frequency and length of supervision, but expect to see technique and frequency/intensity differences appropriate to condition. Similar to an "intent to treat" study, we will include all clinicians as assigned to condition.

C10a. Overview of Hierarchical Modeling Approach

Our analyses for Aim 1 & 2 use Hierarchical Generalized Linear Modeling (HGLM), which has a number of advantages (e.g., allows for nesting; reduces Type I errors from violating assumptions of independence; allows time-varying assessments). We test hypotheses through standard model-building, focusing on creating parsimonious models (Raudenbush et al., 2002; Singer et al., 2003). A null model will establish total variance possible. For longitudinal models (i.e., client outcomes), an "unconditional growth" model will be fit to identify time trends (an unnecessary step for non-growth models [i.e., fidelity]). We will examine the effect of the primary predictor of interest on this base model, and this variable will remain in future models. Finally, a series of control predictors will be fit as follows. Lower level models will be built iteratively in order of theoretical interest and effect size, examining fixed and random effects as indicated. Variables not significantly contributing to the model at $p < .10$, based on likelihood ratio tests and not affecting other coefficients, will be removed. Level-2 predictors will be fit on this level-1 model and will be excluded for the same reasons. In this manner, we create a parsimonious model, including only covariates with explanatory value, which increases power and controls for confounding. Goodness-of-fit will be evaluated using likelihood ratio tests and deviance statistics. Inference will be evaluated relative to $p < .05$.

C10b. Primary Analyses

Aim 1. We explore and describe supervision as provided prior to intervention (SIS, Phase I) using grand means and variability (i.e., over all sessions) and supervisor-level means and variability for frequency and intensity, building on procedures in Garland et al., 2010. This will establish baseline supervision and fidelity, as well as the types of supervision practices used overall and within-supervisor variability in strategy use.

Aim 2; Hypothesis 1a and 1b. Fidelity will be higher for SFM+BR than SFM and higher for SFM than SIS (1a); Client outcomes will be better for SFM+BR than SFM and better for SFM than SIS (1b). To test 1a, we will build a 2-level HGLM predicting fidelity at Phase II. Level-1 variables will include the individual-mean fidelity score for each of the three clients per clinician in Phase II (our primary DV), and client scores on baseline PTSD-RI and SDQ as covariates to control for severity. Level-2 variables will include supervision condition (our primary IV; SFM vs. SFM+BR) and Phase I fidelity. We will explore for and include other covariates (e.g., clinician experience, demographics, attitudes towards EBPs).

Our strategy will: 1) examine the effect of supervision condition on fidelity; 2) statistically control for and examine the impact of important covariates related to fidelity; 3) ensure randomization validity; 4) provide percentage of variance in fidelity accounted for by clinician/client factors and supervision; and 5) allow the inclusion of multiple observations (multiple children per clinician) resulting in precise and reliable estimates of fidelity. Although clinicians are nested within supervisors, a 3-level model is unnecessary because our randomization scheme balances the effect of supervisor characteristics across both conditions.

Analyses to test 1b, whether supervision condition impacts client functioning, will be similarly structured. We will build 3-level longitudinal models, with level-1 variables representing time and child scores, level-2 representing child-level factors, and level-3 representing clinician-related factors (separate models for SDQ & PTSD-RI).These models allow exploration of any significant differences in child functioning due to supervision condition at each time-point (intercepts for BL, 3M, and ET), rates of change in functioning over time (slopes), linear form of change slopes, and variances. Between-group analyses will be used to compare SFM to SFM+BR to examine differences between levels and slopes. We will also conduct within-group analyses, after examining attrition bias (we examine attrition bias for all within-subjects

analyses), comparing SIS to the SFM and SFM+BR using the sample of clinicians retained from Phase I to Phase II..

Aim 3. Hypothesis 2: The impact of supervision condition on client outcomes will be mediated by fidelity. Mediational analyses will employ HGLMs using methods that extend traditional mediational modeling (Baron et al., 1986) to a multi-level longitudinal framework (Cole et al., 2003; Mackinnon, 2008; Stice et al., 2007). H1a and H1b will establish relationships between supervision condition and fidelity, and between supervision condition and child outcomes; additional models will examine the relationship between fidelity and outcomes. Mediation will be determined if the effect of supervision condition on outcomes, controlling for the mediator, is reduced (partial mediation) or eliminated (full mediation); and the sequencing of effects supports the hypothesis that mediation occurs as part of a causal chain. Partial correlation coefficients will estimate effect size. Change in fidelity will be generated as intercept estimates during modeling for H1a. Mediation will be assessed using a difference in coefficients test, estimating the mediated effect and its standard error (Mackinnon, 2008; Sandler et al., in press). Temporal sequencing will be assessed by determining percent of participants with a .5 *SD* improvement in fidelity at 3-mo before a .5 *SD* improvement in functioning at end of treatment, representing a medium effect size (Cohen, 1988). A binomial test will assess if the proportion of participants with a meaningful change in fidelity prior to the moderator is greater than chance.

Aim 4 (Secondary Aim). We will explore the relationship between supervision condition and broader implementation factors and impact on agency-relevant outcomes. We will perform separate analyses on dependent variables (DV) of interest (e.g., supervision alliance, turnover intention, burnout, attitudes towards EBPs, & organizational climate). Mixed between- and within-subjects ANOVAs will test for differences between supervision conditions (between-clinician) and Phase I and Phase II (within-clinician) differences while controlling for important covariates. Estimated marginal means for each phase, stratified by condition, will estimate the impact of condition on change in the DV. Statistical significance will be tested using within-subjects and between-subjects F-tests. Tukey's Honestly Significant Difference test will be used to compare the three supervision groups on mean clinician scores at Phase II 1-year and ET time-points while controlling for family-wise error rates. This analytic approach is analogous to a pre-post (i.e., Phase I-Phase II) design. Although within-subjects analyses are more impacted by attrition, they are also more highly powered to start, such that reduced *N* has less of a negative effect. To examine feasibility and acceptability, assessed via qualitative methods, we will follow procedures used in Drs. Pullmann and Dorsey's prior research using qualitative methods (Dorsey et al., under review; Pullmann et al., 2010). Interviews will be transcribed and coded in the NVivo software program (QSR, 2010) using an integrated grounded theory approach to capture identified categories of interest as well as subcategories or emergent categories. Thematic content analysis, using the constant comparative method, also will be conducted (Patton, 1990; Strauss & Corbin, 1990).

Power. Power analyses were conducted using Optimal Design 2.0 (Spybrook et al., 2009). A primary concern involves the Interclass Correlation Coefficient (ICC) for clinicians, as clinicians are nested within conditions. Higher ICCs result in less statistical power, and few estimates of clinician ICCs for fidelity are available. A review of 20 mental health therapy studies found low clinician ICCs (-0.1 to .06; Baldwin et al., 2011); as did a MST fidelity study (.05; Schoenwald et al., 2008). Because precise estimates are unknown, we examined ICC thresholds within a range of effect sizes (with .05 Type I error rate, power =.80, 3 clients/clinician, 50 clinicians/group). Estimates did not include the effect of covariates that, if moderately correlated with the DV, will increase power. Therefore, our estimates are conservative. H1a is sufficiently powered to detect small effects (.35) with ICCs $\leq$ .07, moderate effects (.50) with ICCs $\leq$ .68, and all large effects. Our test for H1b assumes these same parameters, as well as linear change, and reliability of SDQ = .73 and reliability of PTSD-RI = .9.

For H1b, we have > .80 power on tests of the SDQ with clinician ICC's < .22, and > .80 power on tests of PTSD-RI with ICCs < .23. For H2 (Aim 3), which involves testing for mediation as described above, we have sufficient power for determining the effects of supervision assignment (IV) on both fidelity (mediator) and child outcomes (DV), allowing testing for the mediating effects of fidelity. Tests for H4 are also considered exploratory; however, power analyses indicate sufficient power for these four DVs.

C11. Rationale of Choices Made in the Research Design

Alternative Randomization Strategies. Our team considered randomizing by supervisor to lessen risk of experimental drift, but determined the potential methodological problems (e.g., individual supervisor effects confounded with condition effects) to be a greater risk, especially given our ongoing checks for drift and procedures for re-training. Studies have demonstrated that clinicians can provide distinct but related therapy conditions (e.g., Christensen et al., 2004); therefore, we expect supervisors can also deliver distinct supervision conditions with only one clear, circumscribed, concrete difference (provision of BR or not). Additionally, supervisor-level randomization would severely impact statistical power; with our realistic supervisor sample (*N* = 20) an extremely modest supervisor-level ICC of .02 would result in standard errors nearly twice as large as randomizing by clinician. With a more likely ICC of .1 to .2, standard errors would range from 3.9 to 5.4 times as large. *Alternative Experimental Designs.* We considered alternative designs, including factorial designs isolating particular strategies (e.g., SM only, BR only). However, our goal is to provide recommendations about effective and feasible strategies, and given that supervisors have already had some exposure to gold standard strategies, isolating some is artificial and unlikely to lead to clear recommendations. We also considered a one-phase study, beginning immediately with randomization to SIS, SFM, or SFM+BR in order to lower clinician attrition risk (i.e., shorter study duration). This design precludes establishment of a baseline for CBS current practice (SIS) and reduces power (i.e., no within-subject analyses, fewer clinicians in each group). *Decision not to include observation.* Our team also considered including tape review as one of the Phase II conditions. However, monitoring and behavioral rehearsal strategies are likely more feasible in community settings (Shoenwald et al., 2009), and therefore these strategies received priority. We considered adding a third condition to our current design—supplementing SFM+BR with observation (via tape review) —but adding a third condition reduced power to unacceptable levels and could dilute the potential benefit of adding observation if all strategies are delivered during the current allotted supervision time (e.g.,1-hour). If supervision time was increased for this condition, condition would be confounded by time. In our current design, we are not prescribing (by condition) *or* prohibiting tape review, although we expect it to occur infrequently based on our prior research. If tape review is spontaneously used, we will capture that information with our coding system. Given these tradeoffs and our goal of having effective and feasible recommendations, we believe the current design is the best-suited one for accomplishing study aims.

C12. Hazardous Procedures/Materials: None are planned or expected.

C13. Implications

The proposed research will address a gap in implementation research by testing effective and feasible gold standard supervision strategies that can be consistently and systematically used in community-based settings to support clinician delivery of EBPs. The ultimate goal is to improve mental heath care offered in real world settings, and improve outcomes for children and adolescents. If strategies studied in this proposal are effective and feasible, these strategies can be included in other EBP efforts and will have a substantial public health impact in the area of child and adolescent mental health.