



# Therapist Self-Report of Fidelity to Core Elements of Family Therapy for Adolescent Behavior Problems: Psychometrics of a Pragmatic Quality Indicator Tool

Aaron Hogue<sup>1</sup> · Molly Bobek<sup>1</sup> · Nicole Porter<sup>1</sup> · Alexandra MacLean<sup>1</sup> · Lila Bruynesteyn<sup>1</sup> · Amanda Jensen-Doss<sup>2</sup> · Craig E. Henderson<sup>3</sup>

Accepted: 28 August 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

Therapist-report measures of evidence-based interventions have enormous potential utility as quality indicators in routine care; yet, few such tools have shown strong psychometric properties. This study describes reliability and validity characteristics of a therapist-report measure of family therapy techniques for treating adolescent conduct and substance use problems: Inventory of Therapy Techniques for Core Elements of Family Therapy (ITT-CEFT). Study participants included 31 staff therapists treating 68 adolescent clients in eight community-based mental health and substance use clinics. Therapists submitted ITT-CEFT checklists and companion audio recordings for 189 sessions. The ITT-CEFT contains 13 techniques identified as core elements of three manualized family therapy models that are empirically supported for the target group. Therapists also reported on their use of three motivational interventions, and independent observers coded the submitted recordings. ITT-CEFT factor validity was shown via confirmatory factor analyses of the tool's theoretical structure. Derived modules were: Family Engagement (four items; Cronbach's  $\alpha = .72$ ); Relational Orientation (five items;  $\alpha = .74$ ); and Interactional Change (four items;  $\alpha = .66$ ). Concurrent validity analyses showed fair-to-excellent therapist reliability compared to observer ratings (ICCs range .64–.75); they showed moderate therapist accuracy compared to observer mean scores, reflecting a tendency to overestimate delivery of the techniques. Discriminant validity analyses showed tool differentiation from motivational interventions. Results offer provisional evidence for the feasibility of using the therapist-report ITT-CEFT to anchor quality procedures for family therapy interventions in real-world settings.

**Trial Registration:** The parent clinical trial is registered at [www.ClinicalTrials.gov](http://www.ClinicalTrials.gov), ID: NCT03342872 (registration date: 11.10.17).

**Keywords** Family therapy · Fidelity · Quality indicator · Therapist self-report · Usual care · Adolescent behavior problems

## Quality Indicator: Therapist Report of Fidelity to Evidence-Based Interventions

The behavioral healthcare market is committed to establishing nationwide mandates for increasing quality and accountability in behavioral services delivered in routine settings (Institute of Medicine, 2015). A primary focus of these efforts is promoting the adoption and sustainment of evidence-based interventions (EBIs). Implementation research during the past two decades has convincingly shown that successful delivery of EBIs in mainstream behavioral care requires rigorous quality assurance procedures designed to ensure that interventions are delivered with *fidelity*, that is, in accordance with

✉ Aaron Hogue  
ahogue@toendaddiction.org

<sup>1</sup> Family and Adolescent Clinical Technology & Science, Partnership To End Addiction, 711 Third Avenue, Suite 500, New York, NY 10017, USA

<sup>2</sup> Department of Psychology, University of Miami, Miami, FL, USA

<sup>3</sup> Department of Psychology, Sam Houston State University, Huntsville, TX, USA

research-delineated principles and techniques specific to the given EBI (Frank et al., 2020; Stirman, 2020).

To be effective in promoting EBI delivery in usual care, EBI quality procedures need to be anchored by fidelity metrics that reliably define and assess the nature and merit of the delivered interventions. EBI fidelity metrics can thereby function as quality indicators for behavioral health services (McLeod et al., 2013). Quality indicators in the form of EBI fidelity tools have long been a staple of controlled behavioral research. Unfortunately, the vast majority of fidelity tools and procedures used in controlled research are not pragmatic to deploy in frontline settings. Two main barriers prevent fluid research-to-practice transfer of this valuable technology: Research studies favor observational fidelity methods wherein independent judges evaluate substantial proportions of recorded treatment sessions, a gold standard for assessment rigor (Stirman, 2020) that is beyond reach for resource-challenged treatment systems; and research fidelity tools are typically tailored to the unique training and implementation procedures of the given protocol for which they are companion instruments, making them an inadequate fit for any provider not implementing that protocol (Hogue et al., 2013). Because the pipeline of research-developed EBI fidelity tools is fundamentally inaccessible to the general workforce, quality indicators designed to monitor and evaluate EBI delivery in routine care remain in woefully short supply (Herschell et al., 2020).

This study describes basic psychometric properties of a therapist-report EBI fidelity tool, *Inventory of Therapy Techniques for Core Elements of Family Therapy* (ITT-CEFT). Therapist self-report measures of EBI fidelity have several methodological features that heighten their value in routine care. They are quick, inexpensive, and non-intrusive to use; they efficiently remind and guide clinicians to employ a roster of suitable techniques for a given client; they capture the unique viewpoint of the clinician delivering the interventions; and they can be completed throughout treatment, which facilitates measurement of infrequent but clinically meaningful interventions (Brookman-Frazer et al., 2020; Weersing et al., 2002). Therapist-report measures can also inform quality procedures via data feedback loops of several kinds: as a self-check by therapists to mark their own progress in treating cases; as a supervision aid for trainers to monitor EBI fidelity; and as administrative data for stakeholders and reviewers to evaluate clinician- and agency-level performance (Becker-Haimes et al., 2021; Brookman-Frazer et al., 2020). For these reasons, developing and validating therapist-report tools remains a top priority for advancing EBI implementation in multiple behavioral care sectors (Schoenwald et al., 2011).

## ITT-CEFT Precursor: ITT-ABP, a Multi-Approach Therapist-Report Fidelity Tool

The ITT-CEFT is designed to assess clinician delivery of core family therapy techniques for adolescent behavior problems. Family therapy is an evidence-based approach for most behavioral disorders presented by adolescents in routine care: conduct problems and delinquency (Dopp et al., 2017; McCart & Sheidow, 2016), depression (Weersing et al., 2017), substance misuse (Hogue et al., 2018), and eating disorders (Lock, 2015). Further, systematic reviews (Hogue et al., 2018; McCart & Sheidow, 2016) and meta-analyses (Baldwin et al., 2012; Dopp et al., 2017; Tanner-Smith et al., 2013) suggest that, compared to other evidence-based approaches, family therapy has perhaps the strongest empirical support for treating adolescent conduct and substance use disorders. These are compelling reasons for intensifying efforts to promote delivery of high-fidelity family therapy interventions for adolescent behavior problems in community settings.

The ITT-CEFT is a second-generation EBI fidelity tool whose design and purpose are based on, and serve to elaborate, a precursor fidelity tool: *Inventory of Therapy Techniques for Adolescent Behavior Problems* (ITT-ABP; Hogue et al., 2014a). The ITT-ABP is a post-session therapist-report measure of treatment techniques representing three approaches that each have a substantial base of research support for adolescent conduct and substance use problems (Chorpita et al., 2011; Hogue et al., 2018; McCart & Sheidow, 2016) and are widely endorsed in frontline settings (Cook et al., 2010; Gifford et al., 2012): cognitive-behavioral therapy, family therapy, and motivational interviewing. ITT-ABP items were originally derived from observational fidelity scales validated during controlled trials of manualized treatments, which is considered an advantageous foundation for EBI fidelity tools intended for use in everyday care (Becker-Haimes et al., 2021; Schoenwald et al., 2011).

In particular, the family therapy (FT) scale of the ITT-ABP has shown considerable psychometric strengths when utilized in routine settings. It has shown factor validity and discriminant validity in behavioral services delivered by a diverse clinical workforce operating in both community and hospital-based clinics (Hogue et al., 2014a). It is one of few therapist-report EBI fidelity scales to demonstrate concurrent validity in the form of robust interrater reliability with independent judges (Hogue et al., 2015)—a notable distinction shared with a therapist-report fidelity measure of family-based contingency management for adolescent substance use (Chapman et al., 2013). The FT scale of the ITT-ABP has also supported benchmarking

analyses wherein family therapists practicing in routine care without training in a manualized FT model achieved levels of fidelity to the FT approach equivalent to those achieved by research clinicians in a controlled trial (Hogue et al., 2017a). It has also evidenced predictive validity in community care: Higher scores on the FT scale predicted one-year decreases in adolescent delinquency, externalizing behavior, and substance use in an ethnically diverse sample (Henderson et al., 2019). Interestingly, those prospective FT fidelity effects were evident for clients attending services featuring alternative treatment approaches as well as those featuring FT.

### Derivation of the ITT-CEFT: Empirical Distillation of Core Elements of Family Therapy

Despite the considerable strengths of the FT scale of the ITT-ABP, two foundational limitations in that tool's measurement scope prompted the need to develop a more articulated FT-focused tool, the ITT-CEFT. First, the content of the ITT-ABP's FT scale was derived from an observational adherence measure (Hogue et al., 1998) tethered to a single manualized FT model for adolescent behavior problems, Multidimensional Family Therapy (Liddle, 2016). To develop a fidelity tool that optimally represents the FT approach for adolescent conduct and substance use problems, it is prudent to draw from a cross-section of the several FT models that are empirically supported for these clinical populations. Second, because user burden must be minimized when designing therapist-report tools, instruments like the ITT-ABP that assess multiple treatment approaches (see also Brookman-Frazee et al., 2020; Hurlburt et al., 2010) must reasonably limit the number of items representing each given approach. As a result, the FT scale of the ITT-ABP contains only eight techniques. It was deemed important to develop a single-approach, FT-focused fidelity tool containing a more comprehensive roster of treatment techniques—a roster that more closely approximates a cohort of “necessary and sufficient” (Stirman, 2020) techniques that could flexibly yet pragmatically support FT training and fidelity goals for the community-based clinical workforce (see Regan et al., 2013).

The roster of FT techniques in the newly developed ITT-CEFT is presented in Table 1. The 13 techniques are grouped into three intervention modules: Family Engagement, Relational Orientation, Interactional Change. This roster was derived from a prior distillation process aimed at identifying the *core elements* of FT for adolescent behavior problems, a process detailed in Hogue et al. (2017b). Core elements are specific therapy techniques common to multiple treatment models for a given disorder (Chorpita

**Table 1** Intra-class correlation coefficients\* for reliability between inter-observer ratings and therapist-therapist ratings on the ITT-CEFT

	Inter-observer ICC <sub>1,2</sub>	Therapist- observer ICC <sub>1,2</sub>
<i>ITT-CEFT Total Scale</i>	<b>.84</b>	<b>.74</b>
<i>Family Engagement Scale</i>	<b>.81</b>	<b>.75</b>
Parent Collaboration	.76	.81
Love and Commitment	.66	.54
Parent Ecosystem	.94	.78
Adolescent Goal Collab	.59	.48
<i>Relational Orientation Scale</i>	<b>.74</b>	<b>.64</b>
Relational Focus	.74	.25
Focus on Process	.57	.25
Reframe	.16	.25
Relational Reframe	.44	-.07
Family-Focused Rationale	.26	.37
<i>Interactional Change Scale</i>	<b>.80</b>	<b>.71</b>
Prepares for Future Interactions	.48	-.03
Arranges/Stimulates Dialogue	.56	.58
Coaches and Processes	.76	.77
Teaches Family Skills	.68	.54
<i>Motivational Interventions Scale</i>	<b>.65</b>	<b>.48</b>
Joins with Adolescent	.45	.28
Motivation for Change	.61	.36
Affirm Self-Efficacy	.73	.36

Bold value indicates a summary scale comprising multiple items

N=189 sessions. \*Intraclass correlation coefficients were calculated using the one-way random formula (ICC<sub>1,2</sub>; Shrout & Fleiss, 1979)

& Daleiden, 2009). They are typically identified by (a) specifying the discrete techniques prescribed by approach-congruent treatment manuals validated in research trials and (b) conceptually distilling these techniques into a smaller number of overlapping elements that are core features of each manual. Thus whereas treatment manuals are predominantly complex, uniform, and disorder-specific, distilled core elements are instead granular, flexible, and—to the degree that a given approach (e.g., FT) targets multiple disorders (e.g., adolescent conduct problems, substance use, depression)—potentially transdiagnostic. These are user-centered intervention features (Lyon & Koerner, 2016) that greatly facilitate EBI delivery, and efficient EBI quality procedures, in routine care.

The distillation process identifying a pool of core FT techniques—a pool from which the roster of ITT-CEFT items was ultimately drawn (fully described in Measures)—relied on observational coding of high-fidelity FT treatment sessions. Hogue et al. (2019) sampled 302 sessions from 196 cases treated with one of three models: Multidimensional Family Therapy (Liddle, 2016), Brief Strategic Family Therapy (Szapocznik & Hervis, 2020), or Functional

Family Therapy (Robbins et al., 2016). The sessions were sampled from two efficacy trials and one purveyor-driven training initiative, and all demonstrated strong adherence to their respective manuals based on model-specific fidelity assessments. Hogue et al. used the respective observational fidelity measures of all three models to code each of the 302 sessions. These triangulated fidelity ratings were then analyzed via exploratory followed by confirmatory factor analysis to derive model-shared techniques—that is, commonly observed FT elements expressed in the fidelity blueprint of each model. Notably, as fallout of this empirical distillation process, only two items from the precursor ITT-ABP were eventually retained on the ITT-CEFT.

## Study Context and Specific Aims

The current study investigated the initial psychometric properties of the ITT-CEFT. Data were drawn from 189 sessions held with 68 clients by 31 clinicians practicing at eight mental health and substance use treatment sites. No study site emphasized the FT approach, but each site wanted to increase use of FT techniques. The instrument was introduced to participating therapists and sites as a quality indicator to support high-fidelity delivery of evidence-based FT interventions among adolescent clients. The ITT-CEFT was pitched as a pragmatic quality tool: relevant to clinician and agency goals, low burden to complete, broad applicability across the spectrum of referred youth, based on instruments with strong psychometric properties, and useful for data-driven decision-making (i.e., actionable) (Glasgow & Riley, 2013).

The first study aim was to examine factor validity. As mentioned, the theoretical factor structure of the ITT-CEFT posits three clinical modules undergirding 13 items representing specific FT techniques. ITT-CEFT content was initially informed by a conceptual distillation of core FT elements for adolescent behavior problems (Hogue et al., 2017b) that was subsequently articulated via empirical distillation procedures using observational data from three validated FT fidelity measures (Hogue et al., 2019). Because the ITT-CEFT was developed with a specific theoretical structure, and informed by a previous empirical distillation process (see Measures), we deemed it appropriate to proceed with confirming, rather than exploring, its factor structure. We therefore used confirmatory factor analysis followed by inspection of inter-item correlations to discern whether the tool's empirical factor structure conformed to its theoretical structure.

The second aim was to examine concurrent validity. It is important for therapist-report EBI fidelity to show reasonable concordance with non-participant ratings that are considered the gold standard. We compared ITT-CEFT data to

observational data collected from trained coders who rated session recordings using an observer-report version of the tool. We tested two related dimensions of concurrent validity: reliability and accuracy. Therapist *reliability* refers to the degree to which therapist self-ratings of fidelity covary with observer ratings; this is typically operationalized with inter-rater reliability coefficients. For youth clients, reliability of community clinicians reporting on their own EBI delivery is generally fair-to-poor (e.g., Herschell et al., 2020; Hurlburt et al., 2010); for most such tools, half or more items register below threshold for adequate reliability (e.g., Borntrager et al., 2015; Brookman-Frazee et al., 2020). Interestingly, two therapist-report measures of fidelity to the FT approach (Chapman et al., 2013; Hogue et al., 2015) prove an exception to this rule, having evidenced fairly robust reliability coefficients. Therapist *accuracy* refers to the degree to which mean scores of EBI fidelity (i.e., EBI quantity or dose) based on therapist self-ratings match those based on observer ratings; this is typically operationalized with mean comparisons. Previous research with youth suggests that compared to observers, community clinicians largely overestimate the extent to which they delivered EBIs (e.g., Borntrager et al., 2015; Herschell et al., 2020); this occurs for all approaches, including FT (e.g., Chapman et al., 2013; Hogue et al., 2015). More specifically, therapists tend to over-report both the number (breadth) and extensiveness (depth) of EBI techniques that they have themselves delivered (e.g., Brookman-Frazee et al., 2020; Hurlburt et al., 2010).

The third aim was to examine discriminant validity. We compared mean scores of therapist self-ratings of FT techniques to both therapist self-ratings and observer ratings of three techniques representing motivational interventions: collaborating with the adolescent, increasing client motivation for change, and affirming client self-efficacy. Because motivational interventions of this kind are ubiquitous in behavioral care (Cook et al., 2010), therapist-report fidelity scores for FT techniques should be substantially lower than scores for motivational techniques, especially given that participating sites did not primarily emphasize the FT approach.

## Methods

### Study Participants

Study participants included 31 staff therapists working in community-based mental health and substance use clinics. Therapists (84% self-identified female, 16% male) averaged 30.9 (SD = 8.7) years of age. Self-identified race/ethnicity was 71% White Non-Latinx, 13% Latinx, 7% Black/African-American, 3% Asian, and 6% Other. A total of 94% had a master's level degree and 6% an associate's or bachelor's degree. They averaged 3.7 (SD = 4.4) years of post-degree

therapy experience and 1.8 (SD = 2.9) years of employment at the study clinic. The average caseload size was 31.5 (SD = 23.5) clients across individual, group, and family session formats.

### Study Sites, Study Clients, and Session Participation

Study therapists and their clients were affiliated with eight outpatient behavioral treatment clinics: two were licensed as mental health treatment clinics, four licensed as substance use treatment clinics, and two co-licensed to deliver both mental health and substance use services. Clinics were located in urban ( $n=2$ ) or suburban ( $n=6$ ) locations in various regions of a large northeastern state. Each site prescribed weekly or biweekly single-client treatment sessions for most clients, wherein individual-focused and/or family-focused interventions could be delivered, in addition to group sessions also available on site. None of the sites espoused FT as its primary treatment approach or modality. However, all sites expressed desire to enhance family involvement in agency services and increase routine use of FT techniques during behavioral sessions. Clients ( $n=68$ ) were adolescents referred for outpatient care and their families. Adolescents self-identified as 59% female and 41% male; they averaged 17.3 (SD = 2.1; range 13–21) years of age. Self-identified race/ethnicity was 71% White Non-Latinx, 16% Latinx, 6% Black/African-American, and 7% a different category or multiple categories. Study therapists were asked to document who attended each session for which they submitted study data. For the 189 sessions logged in the current study, 66% included the referred adolescent only, 2% a caregiver only, 31% the adolescent and caregiver conjointly, and 1% the adolescent and someone other than the family member conjointly.

### Participant Recruitment, Data Collection, Participant Attrition, and Session Sampling

Study therapists were recruited to participate in a research study involving collection of therapist-report data on use of FT techniques during behavioral treatment sessions with their adolescent (age 13–21 years) caseloads. Sites and therapists were advised that cataloguing therapist-report data on EBIs on tools such as the ITT-CEFT can reinforce delivery of those interventions in everyday care. Moreover, throughout the study the research team distributed monthly status reports to each therapist that summarized ITT-CEFT data from that therapist across all clients for which data were submitted; summarized data were emailed directly to therapists, who were invited to use the summaries to support clinical tracking and decision-making, independently and/or during supervision. Supervisors also received monthly summaries

containing therapist-blinded ITT-CEFT data averaged across all therapists at the given site.

Prior to the start of the study, therapists completed a three-part, 8-h site training introducing the 13 ITT-CEFT items and 3 supplemental motivational interventions (described in Measures) and building a pragmatic clinical understanding of how they can be implemented in routine care. The training was anchored by didactic instruction during which each technique was described by presenting relevant conceptual theory, precisely defining the given technique, drawing associations with and discriminations from the other twelve techniques, and providing several exemplar therapist statements; didactic instruction was supplemented by presenting mock therapy session video segments illustrating 2–4 techniques in various combinations. Therapists were also given a self-report rating manual (also available to them online) containing one-page descriptions of the ITT-CEFT and motivational intervention items, including exemplar statements of what a therapist might say in session when implementing the given technique. During the study, therapists were asked to submit self-report data and companion audio recordings after sessions for as many clients and sessions as possible, regardless of session composition (i.e., which persons participated in the given session). Therapists submitted their self-report data by completing an online survey powered by Qualtrics with fields for recording session composition information and item scores; session audio recordings were submitted via a secure online upload to protected research archives.

At all sites, all therapists who volunteered were accepted into the study. The recruited sample included 55 therapists who agreed to provide therapist-report data. Of these, 24 (44%) attrited from the study either because they left site employ prior to start of data collection ( $n=2$ ; 4% of the recruited sample) or because they did not submit at least one completed data pair ( $n=22$ ; 40% of the recruited sample). The subgroup of 24 attrited therapists did not differ from the 31 retained therapists (i.e., the study sample) on any of the measured demographics.

Study therapists ultimately submitted 286 pairs of self-report fidelity checklists and companion audio recordings, averaging 9.2 (SD = 8.4) pairs per therapist and 3.9 (SD = 4.1) pairs per client. Coding resource availability permitted us to code a maximum of four session recordings per case. For any case with five or more recordings submitted, four were randomly selected for inclusion in the current study; when this occurred, we prioritized selecting from among those sessions for which therapist-report data indicated that a family member attended. Of the 68 clients for whom session data were submitted, 24 (35%) provided five or more sessions, 4 (6%) provided four, 7 (10%) three, 10 (15%) two, and 23 (34%) one, totaling  $N=189$  sessions.

## Study Measures

*Inventory of Therapy Techniques—Core Elements of Family Therapy (ITT-CEFT)*. The ITT-CEFT is a behavioral treatment quality indicator designed to collect post-session therapist-report data on delivery of core treatment techniques associated with the FT approach for adolescent conduct and substance use problems. The ITT-CEFT consists of three clinical modules containing a total of 13 core FT techniques (see Table 1): *Family Engagement* (4 items): Adolescent Goal Collaboration, Parent Collaboration, Love and Commitment, Parent Ecosystem; *Relational Orientation* (5 items): Relational Focus, Focus on Process, Reframe, Relational Reframe, Family-Focused Rationale; *Interactional Change* (4 items): Prepare for Interactions, Stimulate Dialogue, Coach and Process, Teach Family Skills. The ITT-CEFT operationalizes FT fidelity in the form of extensiveness (i.e., quantity, or dose) scores. Therapists indicate the extensiveness with which each technique was utilized in a just-completed session, based on a 5-point Likert-type scale [0 = *Not at all*, 1 = *A little bit*, 2 = *Moderately*, 3 = *Quite a bit*, 4 = *Extensively*] with the following prompt: “Please indicate how extensively you used each technique in today’s session (i.e., thoroughly and/or frequently)”.

ITT-CEFT items were derived from the respective observational fidelity scales of three empirically supported FT models for adolescent behavior problems: Multidimensional Family Therapy (Liddle, 2016), Brief Strategic Family Therapy (Szapocznik & Hervis, 2020), and Functional Family Therapy (Robbins et al., 2016). In a previous study (Hogue et al., 2019), independent judges coded video recordings of 302 high-fidelity treatment sessions from 196 cases pooled from all three models; each session was coded with all three fidelity scales. These triangulated fidelity ratings were then analyzed to identify model-shared treatment techniques via exploratory factor analyses on half the sample; the identified factors were then validated via confirmatory factor analyses on the remaining half. This empirical distillation process yielded four clinical modules containing a total of 21 specific treatment techniques: Interactional Change (6 techniques), Relational Reframe (7), Adolescent Engagement (4), and Relational Emphasis (4). All 21 items showed fair-to-excellent inter-observer reliability using one-way random intraclass correlation coefficients (ICCs; Shrout & Fleiss, 1979), range 0.54–0.91. All four clinical modules showed strong inter-item correlations within module (i.e., robust internal consistency) using Cronbach’s  $\alpha$ , range 0.67–0.93; as well as weak-to-modest bivariate correlations among the four modules (i.e., robust module differentiation) using Pearson’s  $r$ , range 0.04–0.30.

This observationally defined set of four clinical modules containing 21 core techniques, derived from empirical distillation of high-fidelity FT sessions, was then modified for

therapist-report purposes to create a reduced set of three modules containing 13 techniques. Modifications were enacted for both conceptual and pragmatic reasons. First, several items from the original model-specific fidelity scales that represented caregiver-focused treatment engagement techniques loaded on the Relational Reframe factor during distillation; however, based on clinical coherence and therapist training considerations, it was deemed important for the ITT-CEFT to group caregiver-focused engagement items with adolescent-focused engagement items that loaded on the Adolescent Engagement factor (thereby constituting a *Family Engagement* module). Second, also due to coherence and training considerations, remaining items from the Relational Reframe factor were pooled with items on the Relational Emphasis factor (constituting a *Relational Orientation* module). Third, to reduce the number of scale items in order to minimize reporting burden, several of the original 21 items were either combined with similar other items or eliminated due to fundamental redundancy with other items in the same module, leaving a final total of 13 items contained in 3 factors.

*Motivational interventions*. Therapists were asked to report on their use of three treatment techniques commonly associated with enhancing treatment motivation and commitment to change: Builds a supportive relationship with the adolescent; Explores client concerns about problematic behavior, readiness to change behavior, and optimism about success; Affirms client’s ability to change problematic behavior and praises change efforts. In a previous observational study (Hogue et al., 2015), these items demonstrated solid interrater reliability (ICC = 0.60–0.75). In a previous therapist self-report study, they showed solid factor, convergent, and discriminant validity when grouped with additional items representing cognitive-behavioral techniques (Hogue et al., 2014a). Because these items are intended to serve only as discriminant validity contrasts for ITT-CEFT scales, and not to represent a uniform scale themselves, we do not report on their collective internal consistency. Identically to ITT-CEFT items, therapists indicated the extent to which each motivational intervention was used in a just-completed session, based on a 5-point Likert-type scale: 0 = *Not at all*, 1 = *A little bit*, 2 = *Moderately*, 3 = *Quite a bit*, 4 = *Extensively*.

*Inventory of Therapy Techniques—Core Elements of Family Therapy: Observational Version (ITT-CEFT-O)*. The ITT-CEFT-O contains 16 items: 13 items identical to those on the ITT-CEFT, and 3 identical to the supplemental items describing motivational interventions. The 13 FT items were drawn directly from the previous observational study (Hogue et al., 2019) that was the origin for ITT-CEFT items (psychometric properties described above). The ITT-CEFT-O also contains observational scoring guidelines designed to foster reliable and valid scoring by independent judges.

## Observational Coders and Coding Procedures

Observational coders ( $n = 14$ ) were research personnel consisting of undergraduates and graduates with a bachelor's degree ( $n = 9$ ) and graduates with master's level training in social work, psychology, or a related field ( $n = 5$ ). Observers were trained during weekly virtual meetings over the course of two months using review of the ITT-CEFT-O coding manual, in-group coding and review of practice recordings, and exercises to increase understanding of scale items. Study coding commenced once all observers reached a collective threshold reliability of  $ICC = 0.65$  for the preponderance of items, which required approximately ten practice recordings; thereafter, the group met biweekly for supportive training and monitoring of rater drift until coding was completed. Sessions were scored in their entirety (average about 55 min). Two observers were assigned to score each session; observers were randomly paired with each other across the session sample using a randomized block design (Fleiss, 1981).

## Plan of Analysis

Analyses occurred in four stages to examine key psychometric properties of the ITT-CEFT. In preliminary stage 1 analyses, inter-rater reliability of the ITT-CEFT observational version was calculated for the two observers assigned to each session. Reliability coefficients were generated on all 13 items, the three FT scales (Family Engagement, Relational Orientation, Interactional Change), and the FT total score (i.e., averaged across all items) using the one-way random intraclass correlation coefficient ( $ICC_{1,2}$ ; Shrout & Fleiss, 1979). Once adequate ICCs were established, item scores were averaged across observers to yield a single observer score for each item; item scores were then averaged to calculate a FT total score and three FT scale scores.

Stage 2 analyses assessed factor validity of the ITT-CEFT. First, the theoretical factor structure was tested using confirmatory factor analysis (CFA) on the entire sample of 189 sessions to confirm the fit of the three-factor solution described above. As standard in CFA, we assessed model fit using the model Chi-square statistic and two supplementary fit indices, root mean square error of approximation (RMSEA) and comparative fit index (CFI). RMSEA values of 0.06 and below and CFI values of 0.95 and above indicate strong model fit;  $CFI \geq 0.90$  and  $RMSEA \leq 0.08$  indicate adequate fit (Browne & Cudeck, 1993; McDonald & Ho, 2002). CFA was conducted in Mplus 8.3 (Muthén & Muthén, 1998–2017) and used the sandwich variance estimator to account for the nested structure of the data, specifically, sessions nested within clients, who were nested within therapist, who were then nested within sites (Asparouhov, 2005). Once latent factors (representing clinical modules)

and their constituent items (representing treatment techniques) were confirmed, Cronbach's alpha was calculated for each FT scale as an index of internal consistency, and inter-scale bivariate correlations were calculated to assess the strength of relation between pairs of FT scales (i.e., scale differentiation).

Stage 3 analyses examined concurrent validity, assessing both the reliability and accuracy of therapist reports. To assess therapist reliability, we calculated ICCs comparing therapist ratings to observer ratings. To assess therapist accuracy, we tested for equality of means by comparing therapist ratings to observer ratings on FT total score and the three FT scales using independent samples  $t$ -test. We considered testing for mean differences within a multi-level modeling framework to better account for data nesting; however,  $t$ -tests were chosen for analytic facility and ease of interpretation. To reduce the likelihood of Type I error related to data non-independence in nested data sets (see Wampold & Serlin, 2000), we used an adjusted alpha level of  $p < 0.01$  for significance testing. The multilevel modeling literature consistently indicates that not accounting for nesting deflates standard errors but does not impact expected values, in this case, means (Kreft & DeLeeuw, 1998).

Stage 4 analyses explored discriminant validity via comparison of the FT items to Motivational Intervention (MI) items. Following the procedures describe above, ICCs were calculated to examine reliability of inter-observer scores on an averaged MI total score. Average MI total scores were then compared to mean FT total scores, using both therapist ratings and observer ratings, via a series of paired samples  $t$ -tests. Again, to reduce the likelihood of Type I error related to data non-independence in nested data sets, we used an adjusted alpha level of  $p < 0.01$  for significance testing. Then, bivariate (i.e., Pearson's  $r$ ) correlations were calculated to assess the strength of relation between self-report FT total and MI total scores. These correlations did not take into account the multilevel structure of the data, and therefore, were used descriptively and not interpreted inferentially.

## Results

### Preliminary Analyses: Interrater Reliability of Observer-Report Data

Table 1 contains inter-rater reliability data for the 13 FT items, three scales, and total score. ICCs can be interpreted based on: (a) Cicchetti's (1994) criteria for classifying ICC magnitudes, which are ubiquitous in observational coding research on behavioral interventions: below 0.40 is poor, 0.40–0.59 is fair, 0.60–0.74 is good, and 0.75–1.0 is excellent; and/or (b) Koo and Li's (2016) criteria recommended for behavioral measurement theory more broadly:

below 0.50 is poor, 0.50–0.74 is fair, 0.75–0.90 is good, and 0.91–1.0 is excellent. ICCs ranged from 0.59 to 0.94 for the four Family Engagement items; 0.16 to 0.74 for the five Relational Orientation items; and 0.48 to 0.76 for the four Interactional Change items. ICC for the FT total score was 0.84 (Good/Excellent); ICCs for FT scale scores were Family Engagement = 0.81 (Good/Excellent); Relational Orientation = 0.74 (Fair/Good); Interactional Change = 0.80 (Good/Excellent). These data indicated that the observer scores for the FT total score and all three scale scores were adequately reliable to be used in concurrent and discriminant validity analyses planned for Stages 3 and 4.

### Factor Validity: Latent Structure of Therapist-Report Data

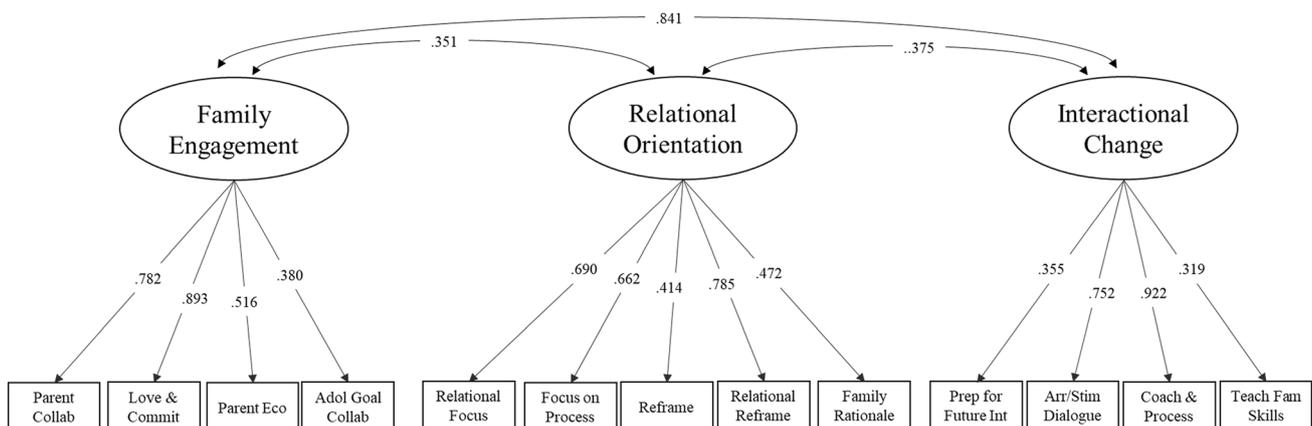
CFA was conducted on the entire sample to confirm the fit of the theoretical three-factor structure of the ITT-CEFT. Model fit was evaluated using chi-square, RMSEA, and CFI. Fit indices for the three-factor solution were:  $\chi^2(62) = 141.36$ ,  $p < 0.001$ ; RMSEA = 0.08 (90% CI 0.06–0.10); and CFI = 0.85, collectively indicating borderline model fit. Figure 1 depicts this solution. When the model was modified by dropping the item with the lowest factor loading (0.32), Teach Family Skills (part of the Interactional Change scale), model fit indices improved to robust fit without further model modification:  $\chi^2(51) = 99.17$ ,  $p < 0.001$ ; RMSEA = 0.07 (90% CI 0.05–0.09); CFI = 0.90. Item-level factor loadings suggested strong factor validity for each module: ranging from 0.38 to 0.90 for Family Engagement items; 0.41 to 0.79 for Relational Orientation items; and 0.34 to 0.93 for Interactional Change items. Although deleting Teach Family Skills from the three-factor solution improved the robustness of overall model fit, we elected

to retain the original solution containing this item, as the item represents a clinically essential behavioral technique in the FT treatment approach and contributes meaningfully to the conceptual integrity of the Interactional Change module; moreover, subsequent psychometric indices for the Interactional Change scale (reported below) suggested that this scale performed on par with the other two scales when retaining the Teach Family Skills item.

Internal consistency for each derived scale was solid as indicated by adequate within-scale inter-item correlations: Cronbach's  $\alpha = 0.72$  for Family Engagement;  $\alpha = 0.74$  for Relational Orientation; and  $\alpha = 0.66$  for Interactional Change. Also, there was meaningful differentiation among scales as indicated by the pattern of between-scale bivariate correlations, with correlations between the latent constructs all falling below  $r < 0.70$  (i.e., non-redundant; Kline, 1979): Family Engagement and Relational Orientation:  $r = 0.43$ ; Family Engagement and Interactional Change:  $r = 0.66$ ; Relational Orientation and Interactional Change:  $r = 0.48$ .

### Concurrent Validity: Therapist-Report Data Compared to Observer-Report Data

Therapist reliability in self-reporting FT technique use was examined via one-way random ICCs comparing therapist-report ratings to gold-standard observer-report ratings on the FT total, three FT scales, and individual items; see Table 1. ICC for the FT total score = 0.74 (Fair/Good); ICCs for the FT scales were: Family Engagement = 0.75 (Good/Excellent), Relational Orientation = 0.64 (Fair/Good), and Interactional Change = 0.71 (Fair/Good). These coefficients demonstrate consistent and fairly robust reliability among therapists estimating their own use of core FT techniques at the total- and module-averaged levels. In contrast, therapist



Note.  $N = 189$  sessions.  $\chi^2(62) = 141.264$ ,  $p < .001$ ; RMSEA = .082; CFI = .852.

**Fig. 1** Confirmatory factor analysis of the ITT-CEFT: Three-factor solution. Note.  $N = 189$  sessions.  $\chi^2(62) = 141.264$ ,  $p < .001$ ; RMSEA = .082; CFI = .852

reliability was highly inconsistent and often poor at the individual item level. ICCs for the 13 FT items ranged from  $-0.07$  to  $0.81$ , with six items falling below  $0.40$  and an additional item below  $0.50$ .

To assess therapist accuracy, a series of independent samples  $t$ -tests were conducted comparing mean scores for therapist-report ratings to observer-report ratings on the FT total score and the three FT scales; see Table 2. Setting an adjusted alpha level of  $0.01$ , results indicate that on average, therapists reported significantly greater scores on the FT total [ $t(376) = -2.89, p = 0.004$ ] and Interactional Change scale [ $t(376) = -3.11, p = 0.002$ ]. No significant difference was found for Family Engagement [ $t(376) = -2.00, p = 0.046$ ] or Relational Orientation [ $t(376) = 0.88, p = 0.388$ ]. These results indicate that therapists were fairly accurate in reporting on the extent to which they used Family Engagement and Relational Orientation techniques, but tended to overestimate their use of Interactional Change techniques and of the entire FT set as a whole.

### Discriminant Validity: FT Extensiveness Compared to MI Extensiveness

As shown in Table 1, inter-observer reliability for the Motivational Intervention scale score (averaging across

3 items) was  $ICC = 0.65$  (Fair/Good), indicating adequate strength to justify use of the scale scores in ITT-CEFT discriminant validity analyses. Discriminant validity of the FT items was examined in three ways. First, as shown in Table 2, for observer-report data, again setting an adjusted alpha level of  $0.01$ , paired samples  $t$ -tests comparing mean MI scale scores to FT scale scores indicated greater use of MI techniques than Family Engagement [ $t(188) = -23.38, p < 0.001$ ], Relational Orientation [ $t(188) = -9.44, p < 0.001$ ], and Interactional Change [ $t(188) = -27.29, p < 0.001$ ]. Second, for therapist-report data, this pattern repeated: MI scale scores were significantly greater than scale scores for Family Engagement [ $t(188) = -15.52, p < 0.001$ ], Relational Orientation [ $t(188) = -19.69, (p < 0.001)$ ], and Interactional Change [ $t(188) = -10.48, p < 0.001$ ]. Third, Pearson's  $r$  coefficients assessing strength of correlation for therapist-report scores indicated that the MI scale was not correlated with the Family Engagement scale ( $r = 0.01$ ), weakly correlated with Interactional Change ( $r = 0.20$ ), and moderately correlated with Relational Orientation ( $r = 0.43$ ). These results show that FT scores were categorically discriminable from MI scores: They were weakly correlated overall, and both therapists and observers consistently rated MI techniques

**Table 2** Descriptive statistics for observer-report ratings and therapist self-report ratings on the ITT-CEFT

	Mean observer ratings			Therapist self-ratings		
	M (SD)	Skewness	Kurtosis	M (SD)	Skewness	Kurtosis
<i>ITT-EF Scale (13 items)</i>	<b>0.69 (0.43)</b>	<b>1.27</b>	<b>1.69</b>	<b>0.83 (0.56)</b>	<b>0.94</b>	<b>0.36</b>
<i>Family Engagement Scale</i>	<b>0.42 (0.52)</b>	<b>1.62</b>	<b>2.49</b>	<b>0.55 (0.71)</b>	<b>1.44</b>	<b>1.12</b>
Parent Collaboration	0.51 (0.91)	1.81	2.29	0.51 (0.96)	1.89	2.73
Love and Commitment	0.20 (0.51)	3.45	13.6	0.59 (1.18)	1.90	2.28
Parent Ecosystem	0.15 (0.63)	5.01	25.81	0.25 (0.65)	3.12	10.61
Adolescent Goal Collab	0.83 (0.89)	0.84	-0.44	0.85 (1.00)	1.08	0.37
<i>Relational Orientation Scale</i>	<b>1.23 (0.55)</b>	<b>0.4</b>	<b>-0.53</b>	<b>1.17 (0.73)</b>	<b>0.56</b>	<b>-0.50</b>
Relational Focus	2.42 (1.09)	-0.14	-1.08	1.40 (1.05)	0.32	-0.76
Focus on Process	1.74 (1.00)	0.05	-0.78	1.48 (1.08)	0.36	-0.59
Reframe	1.51 (0.78)	0.41	-0.12	1.40 (1.04)	0.36	-0.61
Relational Reframe	0.19 (0.46)	2.81	8.31	1.12 (1.21)	0.77	-0.51
Family-Focused Rationale	0.31 (0.54)	1.89	3.21	0.47 (0.83)	1.67	1.71
<i>Interactional Change Scale</i>	<b>0.27 (0.44)</b>	<b>2.22</b>	<b>5.9</b>	<b>0.44 (0.59)</b>	<b>1.65</b>	<b>2.1</b>
Prepare for Future Interactions	0.13 (0.38)	5.04	35.43	0.71 (1.02)	1.38	1.15
Arrange/Stimulate Dialogue	0.20 (0.47)	2.53	6.23	0.30 (0.70)	2.46	5.26
Coach and Process	0.40 (0.85)	2.3	4.56	0.30 (0.76)	2.68	6.63
Teach Family Skills	0.35 (0.76)	2.85	8.7	0.43 (0.84)	2.09	4.08
<i>Motivational Interventions Scale</i>	<b>1.73 (0.63)</b>	<b>0.31</b>	<b>0.21</b>	<b>1.86 (0.93)</b>	<b>0.49</b>	<b>-0.13</b>
Join with Adolescent	2.83 (0.81)	-0.97	1.03	2.25 (1.16)	-0.10	-0.58
Motivation for Change	1.17 (0.92)	0.47	-0.55	1.66 (1.16)	0.48	-0.45
Affirm Self-Efficacy	1.20 (1.00)	0.75	-0.18	1.67 (1.13)	0.36	-0.49

Bold value indicates a summary scale comprising multiple items

N = 189 sessions

as being more prevalent in sessions that FT techniques, as expected for this sample.

## Discussion

Study results show that a new therapist-report post-session measure, Inventory of Therapy Techniques for Core Elements of Family Therapy (ITT-CEFT), demonstrated highly promising psychometric properties in community settings. Evidence of factor validity was substantially, though not fully, supported based on confirmation that all three clinical modules and 12 of 13 individual techniques robustly fit the latent factor structure derived from clinical data provided by 31 community therapists treating 68 clients in eight behavioral treatment sites. Concurrent validity was substantially, though not fully, supported based on adequate-to-strong therapist reliability (i.e., correlations with observer ratings) for all three scales, along with solid evidence of therapist accuracy (i.e., mean score comparability with observer scores) for two of the three scales, though not for the total scale as a whole. And discriminant validity was fully supported, in that the three ITT-CEFT scales had weak correlations with, and were expectably less prevalent than, motivational interventions measured with identical methods.

These promising psychometric properties effectively nominate the ITT-CEFT for provisional duty as a quality indicator (McLeod et al., 2013) for delivery of evidence-based FT techniques for adolescent behavior problems. By design, the tool surmounts two major barriers that commonly interfere with transfer of EBI fidelity measures from research to practice: It focuses on core elements of manualized EBIs rather than a full manual, and it utilizes economical therapist-report methods rather than costly observational methods. Study results indicate that, at least in this initial sample of community-based services, these innovations favoring pragmatic design did not compromise the tool's essential psychometric rigor. Frontline therapists were admirably reliable in reporting on their delivery of all three FT modules: family engagement, relational orientation, and interactional change interventions. Insofar as these three modules represent the fundamental clinical foci of the FT approach for this population (Hogue et al., 2017a, 2017b), the ITT-CEFT appears poised to generate resource-friendly quality metrics for evaluating FT fidelity and fueling data-driven clinical decision-making (Olin et al., 2014).

Study results align with previous research on therapist reliability and accuracy compared to observer ratings. Essentially, indices of reliability address the question, "How much does the length of the ruler change at each measurement?"; in contrast, indices of accuracy address

a validity-related question, "How close is the actual ruler length to the true length?" (see Hallgren, 2012). In the current study, therapists were fairly to excellently reliable at the modular (i.e., aggregated item) level in self-reporting on use of FT techniques with their clients. This extends a trend exhibited in two previous studies with adolescents (Chapman et al., 2013; Hogue et al., 2015) in which therapist-report of FT interventions showed sufficient reliability to support independent use of self-reported data. Also similar to previous research on both FT and other approaches (e.g., Borntrager et al., 2015; Brookman-Frazee et al., 2020), reliability on individual techniques was highly variable, with therapists meeting a benchmark for adequate reliability on only about half of all items. This suggests that for field evaluation purposes, only the total score and module scores, but not individual item scores, be used for quality measurement.

Study therapists were somewhat less strong with regard to self-report accuracy, demonstrating some tendency to overestimate the extent to which they delivered FT interventions. This kind of EBI "dose inflation", a ubiquitous finding in therapist-report fidelity research including FT studies (e.g., Brookman-Frazee et al., 2021; Chapman et al., 2013; Hurlburt et al., 2010), appears deeply rooted in benign reporter biases of several kinds, for example, perceived effort in delivering an intervention, and more inclusive personal framework for evaluating an intervention. One corrective might be to develop inflation-adjustment formulae that can be applied using national or local benchmarks (Hogue et al., 2017a). This might create a workable compromise for utilizing therapist reports of EBI delivery for quality monitoring.

Objectively speaking, therapists in this study did not deliver a high, or even moderate, dose of FT. The therapist-report sample mean score for the FT total scale was 0.83, which falls between the scale anchor values of 0 (*Not at all*) and 1 (*A little bit*); the module score means ranged from 0.44 to 1.17. Low mean scores might be expected given the modest degree of FT allegiance at study baseline—this was definitively not a sample of family therapists. That said, because scale scores were averaged across multiple items, it is entirely possible that in any given session, a given mean score included one or more individual items that received a higher-end score. Clinicians certainly cannot be asked to deliver a full roster of discrete techniques from any governing treatment approach in a single session, in light of prevailing time and client tolerance limits. Indeed, an active therapist can implement one or two interventions very thoroughly during a given session yet still receive a lower-end mean score averaged across multiple items. Another metric for assessing density of EBI delivery in usual care might be tabulating the proportion of sessions in which one (or a few) discrete techniques are scored at or above a midpoint value, indicating the presence of considerable EBI activity (e.g.,

Southam-Gerow et al., 2016), though this strategy requires strong measurement validity at the level of individual scale items.

Contrary to predictions one scale item, Teach Family Skills, was not retained in the final factor analysis solution, though we elected for conceptual reasons to include it in all subsequent analyses. This item is categorically different from the other three items in Interactional Change module in one fundamental aspect: It represents a first-order, rather than second-order, change technique (see Nichols & Schwartz, 1991). In first-order change, family patterns of interactions are altered at the behavioral level only, such that therapists endeavor to bring about observable shifts in sequences of actions. In second-order change, therapists instead target underlying beliefs, premises, or family rules; it is hoped that changes in these latent processes will then prompt observed behavior change (Watzlawick & Weakland, 1977). To take a clinical example: An adolescent and their caregiver may be instructed on using more effective communication strategies to decrease arguing (first-order); or, they may explore and then repair relationship ruptures that have created interpersonal distance and conflict, which would in turn decrease their arguing (second-order). The clinical outcome is the same, but the change processes are fundamentally different (Davey et al., 2011)—which may explain the divergent factor analysis result for Teach Family Skills. Still, retaining this item on the ITT-CEFT helps meet the goal of curating a clinically representative set of first- and second-order FT interventions.

It is critical to emphasize that the 13 techniques on the ITT-CEFT do not represent the full complement of core interventions prescribed by the FT approach for this clinical group. For example, all manualized FT models for adolescent behavior problems (see Henggeler & Schaeffer, 2016; Slesnick et al., 2013) feature some degree of case management interventions—an important category that fell outside the purview of the current study. It is equally important to underscore that core elements by themselves, no matter how lengthy the roster, are not equivalent to manualized treatments. In addition to discrete intervention techniques, treatment manuals invariably articulate principles of treatment coordination—rules for the timing, sequencing, and client- and context-specific targeting of interventions—that constitute the unique parameters and implementation nuances of a given model (Chorpita et al., 2005). Coordination principles determine, for example, how rigidly versus flexibly a therapist should implement model content, as well as the recommended balance between fidelity versus adaptation for individual cases or clinical groups (McHugh et al., 2009). Core elements of EBIs thus cannot supplant full treatment models or be used effectively as “brief” versions of a model.

## Strengths and Limitations

There were several study strengths. The study sampled community therapists operating in unadulterated clinical settings, that is, without benefit of extramural resources of any kind, and they reported on their routine adolescent caseloads. These are conditions of high ecological validity that support generalizability of study findings to real-world practice. Instrument development for the ITT-CEFT was built on a strong foundation: empirically validated observational measures of evidence-based FT models that are widely endorsed for the target group. Analyses followed a multidimensional validity approach (factor, concurrent, discriminant) and observed relatively stringent standards for characterizing reliability and accuracy effects.

There were also numerous study weaknesses. There was a relatively small number of participating sites and therapists, which prevented testing or controlling for site differences and did not supply a nationally representative sample of the usual care workforce. Nearly half of the originally recruited sample dropped from the study because they did not submit any study data; although attrited therapists did not differ from study therapists on any measured variables, this sizable drop off raises concerns about generalizability of study results to the originally recruited group. Collection of recorded sessions was decidedly non-random: Study therapists were asked to record and upload as many sessions as possible, but only a small fraction of convened sessions was submitted, driven by whatever selection biases held sway for a given therapist. These sampling gaps open the door to sampling biases of several kinds (e.g., overrepresentation of therapist-preferred clients and sessions, underrepresentation of clients with erratic attendance or who refused to be recorded) that likewise encroach on study generalizability. Analyses did control for therapist nesting effects but did not investigate individual therapist differences in self-report scores due to lack of power. It was not an aim of this study to isolate variance components for reliability or accuracy data; in future ITT-CEFT studies it would be valuable to learn which therapist and client characteristics account for what proportions of observed variance. With a focus solely on core techniques, the ITT-CEFT does not assess the “contours” of EBI delivery (Schoenwald et al., 2011) defined by the parameters of a given treatment (i.e., service delivery aspects of implementation: to whom, where, how often) and by its prescribed treatment themes and session content (Garland et al., 2010). Asking therapists to judge the (more) readily defined targets and foci of their interventions (i.e., Hogue et al., 2014b), rather than discrete techniques that are often multifaceted and interwoven, sets the measurement bar a notch lower, which might engender improved reliability and accuracy. Another limitation was focus on the extensiveness (i.e., dose) rather than the expertise (i.e., competence) with

which therapists delivered treatment techniques; therapist expertise in implementing specific EBI techniques is highly germane to quality practice but notoriously difficult to judge reliably (Webb et al., 2010).

## Clinical Implications and Future Research

These results provide evidence for the feasibility of using therapist-report EBI fidelity tools to anchor quality procedures for FT interventions in real-world settings. Agency-hired therapists were reasonably concordant with independent observers in judging the extent to which they delivered core FT techniques when treating community referrals for adolescent behavior problems. There is reason to believe that their reliability and accuracy could be even stronger if ongoing training and feedback in documentation of FT delivery were to be incorporated into their routine quality procedures. Such procedures might include training clinicians to self-rate via didactic instruction and in vivo practice guided by experts, along with periodic monitoring of self-report data via supervisor and peer review of ratings (Hogue et al., 2013; Ward et al., 2013). Also, noting the sizable proportion of consented study therapists who submitted no study data, we cautiously suggest that agencies might maximize the completion rate and overall utility of the ITT-CEFT by embedding the tool directly in electronic health records and routinizing ITT-CEFT data summaries in supervision sessions and perhaps staff review procedures.

Future research on the ITT-CEFT should continue to examine model fit and factor generalization to larger and more diverse samples of therapists and clients. An additional step in tool development will be testing predictive validity for targeted client outcomes. Little is currently known about whether EBIs directly influence outcomes in front-line settings, and there is virtually no empirical guidance on which implementation processes are, and are not, essential for producing key effects. Given recent findings that core FT techniques are linked to long-term youth outcomes in usual care (Henderson et al., 2019), such evidence seems worth pursuing.

If the ITT-CEFT continues to demonstrate adequate validity when tested in larger and more diverse therapist and client samples, it would gain standing as a valuable quality tool in efforts to ensure quality implementation of FT in routine behavioral care. Measuring EBI fidelity is becoming increasingly important given system-level healthcare changes that include tracking and incentivizing the use of multilevel quality indicators to assess service delivery (Hoagwood et al., 2013). The ITT-CEFT could serve both prescriptive and evaluative purposes in the difficult task of achieving FT fidelity in front-line settings (see Regan et al., 2013). Therapists could guide family-focused case planning (see Barth et al., 2014) based on the FT techniques contained in the

scale. Supervisors, administrators, and therapists themselves could utilize routine therapist-report documentation of FT fidelity to establish benchmarks for quality performance and provide corrective feedback at the case or clinician level (McLeod et al., 2013).

**Acknowledgements** Preparation of this article was supported by the National Institute on Drug Abuse (R34DA044740; PI: Hogue). The authors gratefully acknowledge the dedicated team of observational coders who made this study possible (in addition to authors MB, AM, LB): Ashli Cleveland, Jenna Cutler, Diana Gonzalez, Cori Hammond, Zachary Kletter, Sosi Korian, Madeline Little, Michelle Martinez, Kelly Roberts, Kennie Swanson, Erica Zakotyansky.

## Declarations

**Conflict of interest** The authors have no conflicts of interest to report.

**Ethical Approval** The study was conducted under approval by the governing Institutional Review Board. Data were collected from March 2019 through March 2020.

## References

- Asparouhov, T. (2005). Sampling weights in latent variable modeling. *Structural Equation Modeling, 12*, 411–434. [https://doi.org/10.1207/s15328007sem1203\\_4](https://doi.org/10.1207/s15328007sem1203_4)
- Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J. Long (Eds.), *Testing Structural Models*. Newbury Park, CA: Sage Publications.
- Baldwin, S. A., Christian, S., Berkeljon, A., & Shadish, W. R. (2012). The effects of family therapies for adolescent delinquency and substance abuse: A meta-analysis. *Journal of Marital and Family Therapy, 38*, 281–304. <https://doi.org/10.1111/j.1752-0606.2011.00248.x>
- Barth, R. P., Kolivoski, K. M., Lindsey, M. A., Lee, B. R., & Collins, K. S. (2014). Translating the common elements approach: Social work's experiences in education, practice, and research. *Journal of Clinical Child & Adolescent Psychology, 43*, 301–311. <https://doi.org/10.1080/15374416.2013.848771>
- Becker-Haimes, E. M., Klein, M. R., McLeod, B. D., Schoenwald, S. K., Dorsey, S., Hogue, A., ... Beidas, R. S. (2021). The TPOCS-self-reported Therapist Intervention Fidelity for Youth (TPOCS-SeRTIFY): A case study of pragmatic measure development. *Implementation Research and Practice. https://doi.org/10.1177/2633489521992553*
- Borntrager, C. F., Chorpita, B. F., Orimoto, T., Love, A., & Mueller, C. W. (2015). Validity of clinician's self-reported practice elements on the monthly treatment and progress summary. *The Journal of Behavioral Health Services & Research, 42*(3), 367–382. <https://doi.org/10.1007/s11414-013-9363-x>
- Brookman-Frazee, L., Stadnick, N. A., Lind, T., Roesch, S., Terrones, L., Barnett, M. L., ... Lau, A. S. (2020). Therapist-Observer Concordance in Ratings of EBP Strategy Delivery: Challenges and Targeted Directions in Pursuing Pragmatic Measurement in Children's Mental Health Services. *Administration and Policy in Mental Health and Mental Health Services Research. https://doi.org/10.1007/s10488-020-01054-x*
- Chapman, J. E., McCart, M. R., Letourneau, E. J., Sheidow, A., & J. (2013). Comparison of youth, caregiver, therapist, trained, and

- treatment expert raters of therapist adherence to a substance abuse treatment protocol. *Journal of Consulting and Clinical Psychology*, 81, 674–680. <https://doi.org/10.1037/a0033021>
- 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 and Clinical Psychology*, 77, 566–579. <https://doi.org/10.1037/a0014565>
- Chorpita, B., Daleiden, E., Ebesutani, C., Young, J., Becker, K., Nakamura, B., et al. (2011). Evidence-based treatments for children and adolescents: An updated review of indicators of efficacy and effectiveness. *Clinical Psychology: Science and Practice*, 18, 154–172. <https://doi.org/10.1111/j.1468-2850.2011.01247.x>
- Chorpita, B. F., Daleiden, E., & Weisz, J. (2005). Identifying and selecting the common elements of evidence based interventions: A distillation and matching model. *Mental Health Services Research*, 7, 5–20. <https://doi.org/10.1007/s11020-005-1962-6>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Cook, J. M., Biyanova, T., Elhai, J., Schnurr, P. P., & Coyne, J. C. (2010). What do psychotherapists really do in practice? An internet study of over 2000 practitioners. *Psychotherapy: Theory, Research, Practice, Training*, 47, 260–267. <https://doi.org/10.1037/a0019788>
- Davey, M., Duncan, T., Kissil, K., Davey, A., & Fish, L. S. (2011). Second-order change in marriage and family therapy: A web-based modified Delphi study. *American Journal of Family Therapy*, 39, 100–111. <https://doi.org/10.1080/01926187.2010.530929>
- Dopp, A. R., Borduin, C. M., White, I., Mark, H., & Kuppens, S. (2017). Family-based treatments for serious juvenile offenders: A multilevel meta-analysis. *Journal of Consulting and Clinical Psychology*, 85, 335. <https://doi.org/10.1037/ccp0000183>
- Fleiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability studies. *Applied Psychological Measurement*, 5, 105–112. <https://doi.org/10.1177/014662168100500115>
- Frank, H. E., Becker-Haimes, E. M., & Kendall, P. C. (2020). Therapist training in evidence-based interventions for mental health: A systematic review of training approaches and outcomes. *Clinical Psychology: Science and Practice*, 27(3), e12330. <https://doi.org/10.1111/cpsp.12330>
- Garland, A. F., Hurlburt, M. S., Brookman-Frazee, L., Taylor, R. M., & Accurso, E. C. (2010). Methodological challenges of characterizing usual care psychotherapeutic practice. *Administration and Policy in Mental Health and Mental Health Services Research*, 37, 208–220. <https://doi.org/10.1007/s10488-009-0237-8>
- Gifford, E. V., Tavakoli, S., Weingardt, K. R., Finney, J. W., Pierson, H. M., et al. (2012). How do components of evidence-based psychological treatment cluster in practice? A survey and cluster analysis. *Journal of Substance Abuse Treatment*, 42, 45–55. <https://doi.org/10.1016/j.jsat.2011.07.008>
- Glasgow, R. E., & Riley, W. T. (2013). Pragmatic measures: What they are and why we need them. *American Journal of Preventive Medicine*, 45(2), 237–243. <https://doi.org/10.1016/j.amepre.2013.03.010>
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8, 23. <https://doi.org/10.20982/tqmp.08.1.p023>
- Henderson, C. E., Hogue, A., & Dauber, S. (2019). Family therapy techniques and one-year clinical outcomes among adolescents in usual care for behavior problems. *Journal of Consulting and Clinical Psychology*, 87, 308–312. <https://doi.org/10.1037/ccp0000376>
- Henggeler, S. W., & Schaeffer, C. M. (2016). Multisystemic therapy: Clinical overview, outcomes, and implementation research. *Family Process*, 55, 514–528. <https://doi.org/10.1111/famp.12232>
- Herschell, A. D., Quetsch, L. B., & Kolko, D. J. (2020). Measuring adherence to key teaching techniques in an evidence-based treatment: A comparison of caregiver, therapist, and behavior observation ratings. *Journal of Emotional and Behavioral Disorders*, 28(2), 92–103. <https://doi.org/10.1177/1063426618821901>
- Hoagwood, K., Atkins, M., & Ialongo, N. (2013). Unpacking the black box of implementation: The next generation of policy, research and practice. *Administration and Policy in Mental Health and Mental Health Services Research*, 40, 451–455. <https://doi.org/10.1007/s10488-013-0512-6>
- Hogue, A., Bobek, M., Dauber, S., Henderson, C. E., McLeod, B. D., & Southam-Gerow, M. A. (2017a). Distilling the core elements of family therapy for adolescent substance use: Conceptual and empirical solutions. *Journal of Child & Adolescent Substance Abuse*, 26, 437–453. <https://doi.org/10.1080/1067828x.2017.1322020>
- Hogue, A., Bobek, M., Dauber, S., Henderson, C. E., McLeod, B. D., & Southam-Gerow, M. A. (2019). Core elements of family therapy for adolescent behavior problems: Empirical distillation of three manualized treatments. *Journal of Clinical Child and Adolescent Psychology*, 48, 29–41. <https://doi.org/10.1080/15374416.2018.1555762>
- Hogue, A., Dauber, S., & Henderson, C. E. (2014a). Therapist self-report of evidence-based practices in usual care for adolescent behavior problems: Factor and construct validity. *Administration and Policy in Mental Health and Mental Health Services Research*, 41, 126–139. <https://doi.org/10.1007/s10488-012-0442-8>
- Hogue, A., Dauber, S., & Henderson, C. E. (2017b). Benchmarking family therapy for adolescent behavior problems in usual care: Fidelity, outcomes, and therapist performance differences. *Administration and Policy in Mental Health and Mental Health Services Research*, 44, 626–641. <https://doi.org/10.1007/s10488-016-0769-7>
- Hogue, A., Dauber, S., Henderson, C. E., & Liddle, H. A. (2014b). Reliability of therapist self-report on treatment targets and focus in family-based intervention. *Administration and Policy in Mental Health and Mental Health Services Research*, 41, 697–705. <https://doi.org/10.1007/s10488-013-0520-6>
- Hogue, A., Dauber, S., Lichvar, E., Bobek, M., & Henderson, C. E. (2015). Validity of therapist self-report ratings of fidelity to evidence-based practices for adolescent behavior problems: Correspondence between therapists and observers. *Administration and Policy in Mental Health and Mental Health Services Research*, 42, 229–243. <https://doi.org/10.1007/s10488-014-0548-2>
- Hogue, A., Henderson, C. E., Becker, S. J., & Knight, D. K. (2018). Evidence base on outpatient behavioral treatments for adolescent substance use, 2014–2017: Outcomes, treatment delivery, and promising horizons. *Journal of Clinical Child and Adolescent Psychology*, 47, 499–526. <https://doi.org/10.1080/15374416.2018.1466307>
- Hogue, A., Liddle, H. A., Rowe, C., Turner, R. M., Dakof, G. A., & LaPann, K. (1998). Treatment adherence and differentiation in individual versus family therapy for adolescent substance abuse. *Journal of Counseling Psychology*, 45, 104–114. <https://doi.org/10.1037/0022-0167.45.1.104>
- Hogue, A., Ozechowski, T., Robbins, M. S., & Waldron, H. B. (2013). Making fidelity an intramural game: Localizing quality assurance procedures to promote sustainability of evidence-based practices in usual care. *Clinical Psychology: Science and Practice*, 20, 60–77. <https://doi.org/10.1111/cpsp.12023>
- Hurlburt, M. S., Garland, A. F., Nguyen, K., & Brookman-Frazee, L. (2010). Child and family therapy process: Concordance of

- therapist and observational perspectives. *Administration and Policy in Mental Health and Mental Health Services Research*, 37, 230–244. <https://doi.org/10.1007/s10488-009-0251-x>
- Institute of Medicine. (2015). *Psychosocial interventions for mental and substance use disorders: A framework for establishing evidence-based standards*. National Academies Press.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15, 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Kreft, I., & DeLeeuw, J. (1998). *Introducing multilevel modeling*. Sage Publications.
- Liddle, H. A. (2016). Multidimensional family therapy: Evidence base for transdiagnostic treatment outcomes, change mechanisms, and implementation in community settings. *Family Process*, 55, 558–576. <https://doi.org/10.1111/famp.12243>
- Lock, J. (2015). An update on evidence-based psychosocial treatments for eating disorders in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, 44, 707–721. <https://doi.org/10.1080/15374416.2014.971458>
- Lyon, A., & Koerner, K. (2016). User-centered design for psychosocial intervention development and implementation. *Clinical Psychology: Science and Practice*, 23, 180–200. <https://doi.org/10.1111/cpsp.12154>
- McCart, M. R., & Sheidow, A. J. (2016). Evidence-based psychosocial treatments for adolescents with disruptive behavior. *Journal of Clinical Child and Adolescent Psychology*, 45, 529–563. <https://doi.org/10.1080/15374416.2016.1146990>
- McHugh, R., Murray, H., & Barlow, D. H. (2009). Balancing fidelity and adaptation in the dissemination of empirically supported treatments: The promise of transdiagnostic interventions. *Behavior Research and Therapy*, 47, 946–953. <https://doi.org/10.1016/j.brat.2009.07.005>
- McLeod, B. D., Southam-Gerow, M. A., Tully, C. B., Rodríguez, A., & Smith, M. M. (2013). Making a case for treatment integrity as a psychosocial treatment quality indicator for youth mental health care. *Clinical Psychology: Science and Practice*, 20, 14–32. <https://doi.org/10.1111/cpsp.12020>
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus User's Guide* (Seventh ed.). Los Angeles, CA: Muthén & Muthén.
- Nichols, M. P., & Schwartz, R. C. (1991). *Family therapy: Concepts and methods* (2nd ed.). Allyn and Bacon.
- Olin, S. S., Kutash, K., Pollock, M., Burns, B. J., Kuppinger, A., Craig, N., ... Hoagwood, K. E. (2014). Developing quality indicators for family support services in community team-based mental health care. *Administration and Policy in Mental Health and Mental Health Services Research*, 41(1), 7–20. <https://doi.org/10.1007/s10488-013-0501-9>
- Regan, J., Daleiden, E. L., & Chorpita, B. F. (2013). Integrity in mental health systems: An expanded framework for managing uncertainty in clinical care. *Clinical Psychology: Science and Practice*, 20, 78–98. <https://doi.org/10.1111/cpsp.12024>
- Robbins, M. S., Alexander, J. F., Turner, C. W., & Hollimon, A. (2016). Evolution of functional family therapy as an evidence-based practice for adolescents with disruptive behavior problems. *Family Process*, 55, 543–557. <https://doi.org/10.1111/famp.12230>
- Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(1), 32–43. <https://doi.org/10.1007/s10488-010-0321-0>
- Shrout, P., & Fleiss, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>
- Slesnick, N., Erdem, G., Bartle-Haring, S., & Brigham, G. S. (2013). Intervention with substance-abusing runaway adolescents and their families: Results of a randomized clinical trial. *Journal of Consulting & Clinical Psychology*, 81, 600–614. <https://doi.org/10.1037/a0033463>
- Southam-Gerow, M. A., McLeod, B. D., Arnold, C. C., Rodríguez, A., Cox, J. R., Reise, S. P., ... Kendall, P. C. (2016). Initial development of a treatment adherence measure for cognitive-behavioral therapy for child anxiety. *Psychological Assessment*, 28, 70–80. <https://doi.org/10.1037/pas0000141>
- Stirman, S. W. (2020). Commentary: Challenges and opportunities in the assessment of fidelity and related constructs. *Administration and Policy in Mental Health and Mental Health Services Research*, 47(6), 932–934. <https://doi.org/10.1007/s10488-020-01069-4>
- Szapocznik, J., & Hervis, O. E. (2020). *Brief strategic family therapy*. American Psychological Association.
- Tanner-Smith, E. E., Wilson, S. J., & Lipsey, M. W. (2013). The comparative effectiveness of outpatient treatment for adolescent substance abuse: A meta-analysis. *Journal of Substance Abuse Treatment*, 44, 145–158. <https://doi.org/10.1016/j.jsat.2012.05.006>
- Wampold, B. E., & Serlin, R. C. (2000). The consequence of ignoring a nested factor on measures of effect size in analysis of variance. *Psychological Methods*, 5, 425. <https://doi.org/10.1037/1082-989x.5.4.425>
- Ward, A. M., Regan, J., Chorpita, B. F., Starace, N., Rodríguez, A., Okamura, K., Daleiden, E. L., Bearman, S., & Weisz, J. R. (2013). Tracking evidence-based practice with youth: Validity of MATCH and standard manual consultation records. *Journal of Clinical Child and Adolescent Psychology*, 42, 44–55. <https://doi.org/10.1080/15374416.2012.700505>
- Watzlawick, P., & Weakland, J. H. (1977). *The interactional view*. Norton.
- Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 78, 200–211. <https://doi.org/10.1037/a0018912>
- Weersing, R. V., Jeffreys, M., Do, M., Schwartz, K., & Bolano, C. (2017). Evidence base update of psychosocial treatments for child and adolescent depression. *Journal of Clinical Child and Adolescent Psychology*, 46, 11–43. <https://doi.org/10.1080/15374416.2016.1220310>
- Weersing, R. V., Weisz, J. R., & Donenberg, G. R. (2002). Development of the therapy procedures checklist: A therapist-report measure of technique use in child and adolescent treatment. *Journal of Clinical Child Psychology*, 31, 168–180. [https://doi.org/10.1207/s15374424jccp3102\\_03](https://doi.org/10.1207/s15374424jccp3102_03)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.