Core Elements of Family Therapy for Adolescent Behavioral Health Problems: Validity Generalization in Community Settings

Aaron Hogue, Molly Bobek, Nicole Porter, Sarah Dauber, Michael A. Southam-Gerow, Bryce D. McLeod & Craig E. Henderson

Published online: 14 Sep 2021.

Submit your article to this journal ☐

View related articles ☐

View Crossmark data ☐

Routledge
Taylor & Francis Group

Check for updates

Core Elements of Family Therapy for Adolescent Behavioral Health Problems: Validity Generalization in Community Settings

Aaron Hogue [a], Molly Bobek[a], Nicole Porter[a], Sarah Dauber[a], Michael A. Southam-Gerow[b], Bryce D. McLeod[b], and Craig E. Henderson[c]

[a]Family and Adolescent Clinical Technology & Science, Partnership to End Addiction; [b]Department of Psychology, Virginia Commonwealth University; [c]Department of Psychology, Sam Houston State University

ABSTRACT

Objective: The core elements of family therapy for adolescent mental health and substance use problems, originally distilled from high-fidelity sessions conducted by expert clinicians, were tested for validity generalization when delivered by community therapists in routine settings.
Method: The study sampled recorded sessions from 161 cases participating in one of three treatment pools: implementation trial of Functional Family Therapy (98 sessions/50 cases/22 therapists), adaptation trial of Multisystemic Therapy (115 sessions/59 cases/2 therapists), and naturalistic trial of non-manualized family therapy in usual care (107 sessions/52 cases/21 therapists). Adolescents were identified as 60% male and 40% female with an average age of 15.4 years; 49% were Latinx, 27% White Non-Latinx, 15% African American, 3% another race/ethnicity, 6% race/ethnicity unknown. Session recordings (n = 320) were randomly selected for each case and coded for 21 discrete family therapy techniques. Archived data of one-year clinical outcomes were gathered.
Results: Confirmatory factor analyses replicated the factor structure from the original distillation study, retaining all four clinically coherent treatment modules comprised of all 21 techniques: *Interactional Change* (ICC = .77, Cronbach's α = .81); *Relational Reframe* (ICC = .75, α = .81); *Adolescent Engagement* (ICC = .72, α = .78); *Relational Emphasis* (ICC = .76, α = .80). Exploratory analyses found that greater use of core techniques predicted symptom improvements in one treatment pool.
Conclusions: Core techniques of family therapy distilled from manualized treatments for adolescent behavioral health problems showed strong evidence of validity generalization, and initial evidence of links to client outcomes, in community settings.

Rationale for Research on Core Elements of Manualized Family Therapy Models

Specific family therapy models are considered evidence-based treatments for most behavioral health disorders presented by adolescents referred to routine behavioral 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, comprehensive reviews (Hogue et al., 2018; McCart & Sheidow, 2016) and meta-analyses (Dopp et al., 2017; Tanner-Smith et al., 2013) show that compared to other research-supported approaches, family therapy models collectively have the strongest evidence base for treating conduct and substance use disorders. Family therapy models also frequently report significant reductions in adolescent internalizing symptoms (described in Hogue & Liddle, 2009). These are compelling reasons for intensifying efforts to promote delivery of evidence-based

family therapy (FT) interventions for adolescent behavioral health problems in community settings. To serve this purpose, a sizable portfolio of manualized FT models stands ready for dissemination, including several models purveyed by corporate entities that supply standardized training and implementation procedures to provider agencies (e.g., Henggeler & Schaeffer, 2016; Robbins et al., 2016).

Despite their exceptional research portfolio, manualized FT models have not been widely adopted in frontline care (Riedinger et al., 2017). To promote high-fidelity delivery, each purveyor-driven FT model contains a compendium of quality assurance procedures anchored by a multicomponent training toolkit, guidelines for ongoing training and observational consultation from model experts, and implementation support and fidelity tracking methods that feed therapy session data back to providers (see Hogue et al., 2013). Such procedures incur substantial financial and resource costs for initial training plus certification

maintenance. In addition, manualized FTs prescribe numerous complex treatment procedures, often with a fixed intervention sequence – features that can inhibit client-centered treatment selection and tailoring practices favored by community clinicians (Chorpita et al., 2005).

Adoption and implementation barriers such as these are common to manualized treatments of many kinds, not just FT models. In response, experts in youth mental health (e.g., Chorpita et al., 2007; Van Der Pol et al., 2019; Weisz et al., 2011) advocate a complementary strategy to manualized models: focus on core elements of evidence-based treatments. Core elements are specific therapy techniques that are common to multiple treatments for a given disorder (Chorpita & Daleiden, 2009). They are typically identified by (a) specifying the discrete techniques prescribed by approach-congruent treatment manuals validated in research trials and (b) 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., cognitive-behavioral therapy) targets multiple disorders (e.g., childhood anxiety, depression, and conduct problems) – potentially transdiagnostic. These are user-centered intervention features (Lyon & Koerner, 2016) that may help solve vexing barriers associated with implementing treatment manuals in routine care (McHugh & Barlow, 2010). For example, core elements of cognitive-behavioral therapy for childhood depression include activity selection, cognitive restructuring, and coping skills training (see Chorpita & Daleiden, 2009).

It is believed that core elements can complement dissemination of treatment manuals by enhancing the basic clinical competencies of community practitioners (Weisz et al., 2017) as well as co-exist with manualized models to provide a range of treatment planning options for clinicians operating in busy treatment marketplaces (Southam-Gerow et al., 2014). Research support for core elements is rapidly progressing. Randomized trials have reported long-term effectiveness for core elements aimed at treating a broad set of childhood disorders (Chorpita et al., 2017, 2013) as well as enduring training effects among community agencies that adopt core elements (Weisz et al., 2018). However, as with manualized treatments, the robustness of core element interventions in frontline settings may well depend on the type and intensity of implementation support rendered (Weisz et al., 2020). Behavioral treatment research is just beginning to discover the benefits as well as limits of the core elements approach.

Precursor Study and Contributions of the Current Study

This study extends the work of a previous study that distilled the core FT elements from three manualized FT models rooted in a systemic family therapy approach (see Minuchin & Fishman, 1981) that focuses on changing family relationships as well as relations between family members and extrafamilial systems. Hogue et al. (2019) sampled 302 therapy 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). These sessions were termed "gold standard": They 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 analyzed to derive model-shared FT techniques via exploratory factor analyses on a randomly selected half-sample; the derived factors were then validated via confirmatory factor analyses on the remaining half. This empirical distillation process identified four clinically coherent factors (i.e., core modules) consisting of 21 discrete treatment techniques (i.e., core elements) with strong internal consistency: *Interactional Change* (6 techniques; Cronbach's α = .93), *Relational Reframe* (7 techniques; α = .79), *Adolescent Engagement* (4 techniques; α = .68), and *Relational Emphasis* (4 techniques; α = .67). Each module contained techniques from at least two of the three FT models, providing evidence of module commonality across manualized models.

The current study examined these same core FT modules and elements, distilled from highly trained clinicians, in the context of FT delivered by community therapists in routine settings. Whereas research on FT for adolescent behavioral health problems has relied heavily on controlled efficacy or effectiveness designs, two recent studies demonstrated that community therapists can deliver FT techniques in routine adolescent care with fidelity and outcome effects comparable to that shown in efficacy trials (Hogue, Dauber et al., 2017; Hogue et al., 2015). Findings like these, indicating that community clinicians are capable of delivering effective FT without using manualized models, buoy efforts to identify and test core FT elements. Even so, fidelity and outcomes for any evidence-based approach are expected to vary significantly among therapists attempting to implement the approach in frontline care (Deighton et al., 2016; Lau et al., 2017). Research is needed to determine the potential for the structure of

the core FT modules and elements to generalize across FT approaches and settings, such that the specific treatment techniques co-occur in similar fashion when delivered by community clinicians in routine care.

For exploratory purposes, this study also examined links between core FT technique use and client outcomes. If the core FT techniques distilled in these community samples were also linked to client gains, this could further underscore their clinical value. It is important to state that meta-analyses across multiple treatment approaches and populations generally yield modest or negligible technique-outcome effects (see Collyer et al., 2019 for youth studies; Webb et al., 2010 for all age groups). Yet it is also true that studies of manualized FTs for adolescent behavioral health problems often find that greater use of model-specific techniques predicts better outcomes on a variety of clinical indicators (e.g., Gillespie et al., 2017; Hogue et al., 2006, 2008; M. Robbins et al., 2011). Again, those results derived from research on manualized FT models in controlled settings. It remains of interest whether FT technique-outcome links established in the lab can hold form in the more turbulent field of community practice (see Henderson et al., 2019).

Study Hypotheses

The main hypothesis of the current study pertained to validity generalization, defined as the extent to which the psychometric properties of a tool or construct generalize across settings and populations that are different from the ones in which it was originally validated (McLeod et al., 2013). We hypothesized that the four-factor structure of core FT elements for adolescent behavior problems, originally derived using gold-standard samples from controlled settings (Hogue et al., 2019), would replicate in diverse community contexts. We selected 320 sessions from three distinct community treatment pools: an implementation trial of a manualized FT, a protocol adaptation trial of a different manualized FT, and a naturalistic trial of therapists delivering non-manualized FT as the routine standard of care. For all three pools, the parent trial reported strong therapist adherence to the respective FT approach and improved client outcomes at one-year follow-up (as described in Study Sample). Selected therapy sessions were observationally coded for the presence of all 21 techniques from the original core FT element distillation and then tested for fit with the original four-factor solution of core FT modules (see above; Hogue et al., 2019).

On an exploratory basis, we also examined whether higher scores on a composite scale of factor-derived core FT techniques would predict better clinical outcomes across one-year follow-up. We tested technique-outcome links separately for each pool in order to avoid making assumptions about homogeneity in such across pools (i.e., that core FT techniques have similar predictive values in each pool) and, given the exploratory nature of these analyses, to avoid masking links that might obtain in only one or two pools. We focused on outcomes in three domains that were primary reasons for referral in all three pools: externalizing problems, internalizing problems, and substance use. For each pool we analyzed only those outcomes that demonstrated significant change in the selected cases (as there is no possibility to detect technique-outcome effects in the absence of change on the given outcome); this amounted to two of three candidate outcomes per pool (see Measures).

Method

The study was conducted under approval by the governing Institutional Review Board.

Study Sample: Three Community Treatment Pools

This study sampled 320 recorded therapy sessions and clinical outcome data from 161 cases treated by 45 therapists participating in one of three randomized trials testing delivery of FT for adolescent behavior problems in community settings. Adolescents were identified as 60% male and 40% female with an average age of 15.4 years (SD = 1.7); 49% were Latinx, 27% White Non-Latinx, 15% were African American, 3% another race/ethnicity, 6% race/ethnicity unknown. Therapists were identified as 71% female and 39% male; 58% were Latinx, 27% White Non-Latinx, 15% African American. Importantly, therapists in all three pools were standard workforce (versus research-hired) clinicians treating cases from routine clinic referral streams. Outcome data were collected at baseline, 6-, and 12-month follow-up in all three pools.

One pool, an implementation trial of Functional Family Therapy (FFT: Robbins et al., 2016), contributed 98 sessions from 50 cases (70% male and 30% female; average age 15.1 years; 76% Latinx, 16% African American, 6% White Non-Latinx, 2% unknown) treated by 22 therapists (15 female and 7 male; 10 Latinx, 9 White Non-Latinx, 3 African American). The parent FFT trial for this pool (Robbins et al., 2019) occurred in California and examined observation-based supervision versus conventional supervision in the FFT model across eight community agencies staffed by therapists who had received FFT certification from model experts;

the current study drew sessions from both trial conditions. The average number of sessions per case was 14.3 (SD = 3.5).

A second pool, an adaptation trial of Multisystemic Therapy (MST: Henggeler & Schaeffer, 2016), contributed 115 sessions from 59 cases (59% male and 41% female; average age 15.1 years; 66% White Non-Latinx, 14% African American, 5% Latinx, 5% another race/ethnicity, 10% unknown) treated by 2 therapists (1 female and 1 male; 2 African American). The parent trial (Sheidow et al., 2020) occurred in South Carolina and showed the fidelity and efficacy of an adapted version of MST tailored to adolescents with co-occurring substance use and behavior problems compared to usual care in a single community clinic. Study therapists were trained and supervised by an MST expert. Sessions occurred on an approximately weekly basis; treatment duration per case averaged 9.7 months (SD = 4.5).

A third pool, a naturalistic trial of non-manualized FT in usual care (UC-FT), contributed 107 sessions from 52 cases (50% female and 50% male; average age 15.8. years; 73% Latinx, 15% African American, 2% White Non-Latinx, 4% another race/ethnicity, 6% unknown) treated by 21 therapists (16 female and 5 male; 16 Latinx, 3 White Non-Latinx, 2 African American). The parent trial (Hogue et al., 2015) occurred in New York and showed superior outcomes for adolescents with behavior problems in one clinic that featured FT as the routine standard of care versus five clinics that featured various alternative treatment approaches. The UC-FT condition was shown to have strong adherence to the FT approach (Hogue, Dauber et al., 2017). The average number of sessions per case was 14.1 (SD = 11.8).

Session Sample Selection Procedures

To select a reduced but still representative sample of recorded (audio or video) sessions from each study case, treatment duration was divided into three generic sampling phases: Phase 1 (sessions 1–2), Phase 2 (sessions 3–6), Phase 3 (sessions 7–20). Sessions later than 20 were excluded from randomization, as these represent an unusual length for FT in routine care, unless only sessions later than 20 were available; in these few instances (n = 9) the earliest session available was used for Phase 3. One session was randomly selected for coding from each sampling phase for which at least one session had occurred. The final session for each case was excluded to avoid selection of termination-focused sessions that may preclude use of most FT techniques. A percentage of initially selected recordings were over 75 minutes long; in this circumstance (n = 31), a shorter replacement tape within the same phase was

randomly selected, or if a replacement tape was not available, only the first 60 minutes were coded. Fourteen percent of the sample (n = 22) had sessions in all three phases, 68% (n = 109) had sessions in two phases only (usually Phases 2 and 3), and 15% (n = 24) had sessions in one phase only (usually Phase 3). Overall there were 99 recordings in Phase 1, 103 in Phase 2, and 118 in Phase 3, for a total of 320 recordings. Adolescents and caregivers appeared together in 74% of sessions, with 18% of sessions containing only adolescents and 8% only caregivers.

Study Measures: Core FT Techniques

Therapist Behavior Rating Scale: Core Elements of Family Therapy (TBRS-CEFT)

The TBRS-CEFT (Bobek et al., 2018) is an observational rating scale containing 21 items that represent the core elements of evidence-based FT for adolescent behavioral health disorders. The items were empirically distilled from three research-supported manualized FT models using observational fidelity coding methods (Hogue et al., 2019) and constitute a coherent set of specific techniques that are common across models (Hogue, Bobek et al., 2017). The techniques, listed in Table 1, cluster into four modules: *Interactional Change* (6 techniques): interventions in which therapists allow or prompt family members to interact with one another naturally in order to assess family dynamics, and also, direct in-session interactions among members in order to promote more effective ways of relating; *Relational Reframe* (7 items): efforts to transform symptom-focused and/or adolescent-focused perceptions of clinical problems into a new understanding of those problems as being fundamentally relational, thereby motivating families to pursue changes in family relationships as the primary clinical solution; *Adolescent Engagement* (4 items): interventions in which therapists join with adolescents by seeking their unique points of view, and also, foster treatment engagement by presenting family therapy as an opportunity to address personally meaningful issues within and outside the family; and *Relational Emphasis* (4 items): interventions that focus on the family as a whole, assessing systemic attributions and processes and intervening to improve overall family functioning. The scale measures the extensiveness (i.e., thoroughness and/or frequency) with which each technique was used in the observed session, based on a 5-point Likert-type scale: 1 = *Not at all*, 2 = *A little bit*, 3 = *Moderately*, 4 = *Considerably*, 5 = *Extensively*. The original psychometric study (Hogue et al., 2019) showed fair-to-excellent interrater reliability for each item using one-way random intraclass correlation coefficients (ICC; Shrout & Fleiss,

Table 1. Results of confirmatory factor analyses on 21 core family therapy technique items.

	Standardized Factor Loadings				Correlated Residuals
	IC	RR	AE	RE	
1. Moves close to direct restructuring maneuvers	.876				
2. Coaches interactions in session	.702				6, 3
3. Stimulates dialogues/Directs enactments	.701				2
4. Remains decentralized after enactment emerges	.674				
5. Focuses on present interactions	.642				
6. Conducts in-session exercises for new behaviors	.391				2
7. Provides information/guidance for new skill		.854			13, 8
8. Utilizes meaning-change interventions (reframe)		.719			9, 7
9. Offers more positive view of problems/family		.686			8, 12, 11
10. Maintains a relational focus		.643			
11. Develops relational reframe for youth problems		.637			9
12. Targets adult participants for change		.480			9
13. Provides a family-focused rationale for change		.330			7
14. Joins with children/adolescents			.828		
15. Targets interventions toward youth			.779		
16. Supports adolescent investment in therapy			.664		
17. Explores adolescent ecosystem			.511		
18. Asks clarifying questions/Focuses on process				.787	
19. Enhances family attachment/communication				.774	
20. Gathers information on relationship functions				.718	
21. Connects with all family members				.594	

N = 320 sessions across all three study pools: FFT, MST, UC-FT. IC = Interactional Change, RR = Relational Reframe, AE = Adolescent Engagement, RE = Relational Emphasis.

[a]Indicates which error residuals were allowed to correlate with each item in the final factor models. Decisions about allowing residuals to correlate were made based on model modification indices and were predominantly due to overlap in item content. For example, the error term for item 2 was allowed to correlate with the error term for item 3 as both items focus primarily on therapist efforts to facilitate in-session interactions between family members.

1979), range .54 – .91; strong inter-item correlations within module (i.e., internal consistency) using Cronbach's α, range .67 – .93; and weak-to-modest average correlations among the four modules (i.e., strong module differentiation) using Pearson's r, range .04 – .30.

Study Measures: Client Outcomes

This study selected archived clinical outcomes from the parent trials that represented primary reasons for referral in the trials and showed significant change over 1 year for study cases. All client outcome data were collected at baseline, 6-month follow-up, and 1-year follow-up. Caregiver-reported *externalizing problems* (delinquency, aggression) and youth-reported *internalizing problems* (anxiety/depression, social withdrawal, somatic complaints) were measured by the *Child Behavior Checklist* and *Youth Self-Report* respectively (Achenbach & Rescorla, 2001), which are parallel measures of youth behavior problems supported by extensive evidence of reliability, validity, and clinical utility. *Substance Use* was measured with urine drug assays testing for a range of commonly used illegal substances that indicated either absence or presence of a given substance; this study collapsed all results for a given assay into a dichotomous variable: Any versus No substance present. Externalizing problems showed significant change in all three pools (FFT, MST, UC-FT); internalizing problems showed change in the FFT and UC-FT

pools; and substance use showed change in the MST pool only (this was the only pool for which the parent trial stipulated elevated substance use as a study eligibility requirement).

Observational Coding Procedures and Raters

Raters scoring core FT technique use on the TBRS-CEFT were trained during twice-weekly meetings via review of the scoring manual, in-group coding practice, and exercises to increase understanding of scale items. Study coding commenced once raters reached a collective reliability threshold of $ICC_{(1,2)}$ = .65 for 80% of items; reliability was monitored thereafter. Sessions were independently coded in their entirety by two raters randomly assigned to sessions in pairs according to a balanced incomplete randomized block design (Fleiss, 1981). There were 11 raters (10 female, 1 male): 8 White Non-Latinx, 3 Latinx. Six had Master's degrees and three had Bachelor's.

Plan of Analysis

Study analyses occurred in three stages. In Stage 1, interrater reliability statistics for each of the 21 FT technique items were calculated on composite variables that combined data across the three community pools. Interrater reliability was calculated using the one-way random intraclass correlation coefficient (ICC; Shrout &

Fleiss, 1979). Prior to Stage 2 analyses, item scores were averaged across both raters to yield a single score for each technique.

In Stage 2, confirmatory factor analysis (CFA) with correlated factors was conducted on the entire sample of 320 sessions to confirm the fit of the four-factor solution originally derived by Hogue et al. (2019; described above). First, preliminary CFA models were estimated for each factor separately to achieve adequate fit for individual factors, before attempting to fit the full four-factor model. Modification indices were examined to guide theoretically based model adjustments needed to maximize model fit. Model parameters were estimated using maximum likelihood estimation based on conceptualizing the 5-point observational coding scale (TBRS-CEFT) as a continuous scale (Li, 2016), which aligns with how raters were trained to implement the scale when scoring sessions. Model fit was assessed using the model chi-square statistic and two supplementary fit indices, RMSEA and CFI. RMSEA values of .06 and below, and CFI above .95, indicate strong model fit, and CFI ≥ .90 and RMSEA ≤ .08 indicate adequate fit (Browne & Cudeck, 1993; McDonald & Ho, 2002). CFA was conducted in Mplus 7.31 (Muthén & Muthén, 1998/2017). The sandwich variance estimator, which produces corrected standard errors in the presence of nested data (Asparouhov, 2005), was used to account for nesting of sessions within therapists. Respective nest sizes were not large enough to account also for nesting of sessions within cases or therapists within pools.

In Exploratory Stage 3, latent growth curve (LGC) modeling (Duncan et al., 1999) was used to examine FT technique-outcome associations for 152 cases (9 cases included in the CFA did not have complete outcome data on any selected measure and thus were excluded). Scores for each of the 21 core FT technique items were averaged across all observed sessions for each case, yielding a FT Total score that was used as the predictor in all analyses. We favored this unit weighting method over the equally valid option of saving derived factor scores as predictors (see Nunnally & Bernstein, 1994) because the unit of CFA analysis was treatment session, whereas the unit of LGC analysis was case. Each outcome (i.e., dependent variable: DV) was analyzed separately for each pool for reasons described in the Study Hypotheses section. Preliminary one-way analyses of variance tested for pool differences in FT Total score. Also, preliminary analyses examined whether adolescent race/ethnicity, sex, or age was associated with linear change in each DV; Sex was significantly associated with youth-report externalizing symptoms and was subsequently retained as a covariate in technique-outcome analyses for that DV. Missing data were handled with robust maximum likelihood estimation under the assumption that data were missing at random (Little & Rubin, 2002) as was reported in each parent trial. LGC was conducted using Mplus 7.4 (Muthén & Muthén, 1998/2017). To control for cases nesting within therapists we used the sandwich variance estimator in Mplus (Diggle et al., 2002). As expected (because outcomes were selected based on showing change across follow-up for study cases), the unconditional model for each DV demonstrated a significant decline in symptoms over time. For each technique-outcome analysis we calculated the slope coefficient, standard error (SE), and pseudo-z score (calculated by dividing the model-estimated coefficient by its standard error). For effect size estimates, we used β coefficients from LGC models estimating fully standardized effects. Specifically, β indexes unit change in the outcome variable for one standard deviation change in slope of the TBRS-CEFT score (as described in Muthén & Muthén, 1998/2017).

Results

Stage 1. Core FT Techniques and Modules: Interrater Reliability

One-way random ICCs were calculated for each core FT technique item (listed in Table 1) to examine interrater reliability. ICCs can be interpreted based on: (a) Cicchetti's (1994) criteria for classifying ICC magnitudes, criteria that are ubiquitous in observational coding research on behavioral interventions: below .40 is poor, .40–.59 is fair, .60–.74 is good, and .75–1.0 is excellent; and/or (b) Koo and Li's (2016) criteria recommended for behavioral measurement theory more broadly: below .50 is poor, .50–.75 is fair, .75–.90 is good, and .90–1.0 is excellent. ICCs ranged .55 to .74 for the six Interactional Change items; .48 to .69 for the seven Relational Reframe items; .46 to .72 for the four Adolescent Engagement items; and .46 to .70 for the four Relational Emphasis items. Of the 21 total items, 3 (Maintains a relational focus, Supports adolescent investment in therapy, Connects with all family members) registered ICCs below .50; all 21 items were above .40. These data indicated that the set of 21 item-level scores was adequately reliable for Stage 2 factor analyses. ICCs for the mean-level module scores were in the good-to-excellent range: Interactional Change = .77; Relational Reframe = .75; Adolescent Engagement = .72; Relational Emphasis = .76. These data indicated that all module scores were adequately reliable for relevant Stage 2 analyses. These ICC magnitudes and distributions closely approximate those reported in other observational coding studies of FT for adolescent behavioral health problems (e.g., Gillespie et al., 2017; Hogue et al., 2008; M. Robbins et al., 2011).

Stage 2. Core FT Modules: Confirmatory Factor Analysis and Psychometrics

CFA was conducted on the entire sample to confirm the fit of the four-factor solution originally derived by Hogue et al. (2019). Model fit for each module was evaluated using chi-square, RMSEA, and CFI. Fit indices for Interactional Change were: χ^2 (7) = 8.90, p= .25; RMSEA = .03 (90% CI: .00 – .08); CFI = .99. Fit indices for Relational Reframe were: χ^2 (9) = 48.28, p= .00; RMSEA = .12 (90% CI: .09 – .15); CFI = .98. Fit indices for Adolescent Engagement were: χ^2 (2) = 5.82, p= .05; RMSEA = .08 (90% CFI = .00–.15); CFI = .99. Fit indices for Relational Emphasis were: χ^2 (2) = 4.80, p= .10; RMSEA = .07 (90% CI: .00 – .15); CFI = .99. Evaluation of these fit indices indicated that model fit was adequate for each of the four modules individually. As depicted in Table 1, for two of the derived factors (Interactional Change, Relational Reframe), error residuals were allowed to correlate among several within-factor items; decisions about allowing residuals to correlate were made due to overlap in the given item content and were supported by modification indices.

Item-level factor loadings (see Table 2) suggested strong factor validity for each module. Factor loadings ranged from .39 to .88 for Interactional Change items; .33 to .85 for Relational Reframe items; .51 to .83 for Adolescent Engagement items; and .59 to .79 for Relational Emphasis items. Note that model fit indices for Relational Reframe were somewhat weaker than those for other factors, and that two Relational Reframe items fell below the factor loading threshold of .50 (registering .33 and .48) conventionally observed for item retention on a given factor. However, we elected to retain those two items on the factor for related three reasons: Both items contribute meaningfully to the conceptual integrity of the factor; both load more highly on this factor than any other; and by retaining them, we preserve exact factor replication from the previous distillation sample for the remaining Stage 2 and Stage 3 analyses. For identical reasons, we elected to retain one item on the Interactional Change factor whose loading was .39.

When combined into a single CFA model, the full four-factor model failed to converge; this also occurred in the original Hogue et al. study. Internal consistency for each derived module was robust as indicated by strong inter-item correlations within module: Cronbach's α = .81 for Interactional Change; α = .81 for Relational Reframe; α = .78 for Adolescent Engagement; and α = .80 for Relational Emphasis. Also, meaningful differentiation among modules was indicated by the overall pattern of bivariate correlations between modules, wherein each correlation was r < .70 (i.e., non-redundant; Kline, 1979): Interactional Change and Relational Reframe: Pearson's r = .65; Interactional Change and Adolescent Engagement: r = .13; Interactional Change and Relational Emphasis: r = .44; Relational Reframe and Adolescent Engagement: r = .05; Relational Reframe and Relational Emphasis: r = .61; Adolescent Engagement and Relational Emphasis: r = .18.

Exploratory Stage 3. Core FT Technique-Outcome Links: Latent Growth Curve Modeling

Preliminary omnibus testing for mean differences in FT Total score by community pool revealed a significant difference [F(2, 152) = 39.30; p < .001]. A series of post-hoc independent samples t-tests revealed that FT Total score was higher in the FFT pool (M = 2.7, SD = .44) compared to both the UC-FT pool [M = 2.3, SD = .43; t(100) = −5.21; p < .001] and MST pool [M = 2.1, SD = .25; t(101) = 9.34; p < .001]. FT Total score was higher in the UC-FT pool compared to the MST pool [t(103) = 3.03; p = .003]. Results of LGC models in which FT Total score predicted client outcomes across one-year follow-up are presented in Table 3; we report statistically significant results here. The β coefficients used as effect sizes can be interpreted as follows: Small = .10, Medium = .30, Large = .50 (Cohen, 1988). There was a significant association between FT Total score and two outcomes, both within the MST pool. Greater use of core FT techniques in therapy sessions predicted over-time decreases in adolescent externalizing problems (slope coefficient = −3.46, SE = .63, pseudo-z = −5.49, p< .001, β = .21) and increasingly larger proportions of youth abstaining from substance use (slope coefficient = −.64, SE = .15, pseudo-z = 4.13, p < .001, β = .20).

Discussion

Study findings indicate that the core treatment techniques of empirically supported FT models for adolescent behavioral health problems – originally distilled from manualized treatments delivered by highly trained clinicians in controlled settings – co-occurred in similar fashion in FT treatment delivered by community therapists in routine practice. In support of validity generalization, the original set of core FT elements and their module groupings – Interactional Change, Relational Reframe, Adolescent Engagement, Relational Emphasis – were confirmed to be operational across three community-based FT pools. In exploratory analyses, greater use of core FT elements was associated with reduced externalizing problems and substance use in one of the three pools.

The main study finding was that the full set of core FT elements and modules from the original distillation of gold-standard sessions (Hogue et al., 2019) was replicated in

Table 2. Outcome variable descriptive statistics and model estimated means by sample pool.

	N	Descriptive Statistics	Model Estimated	
		Mean (SD)	Mean	SE
FFT Pool				
Internalizing problems				
Baseline	50	54.02 (13.02)	53.75	3.59
6-month	40	52.21 (14.60)	52.01	
12-month	43	50.00 (12.12)	50.27	
Externalizing problems				
Baseline	42	66.24 (17.92)	65.31	
6-month	39	61.14 (16.31)	61.79	3.45
12-month	42	60.99 (20.60)	58.27	
UC-FT Pool				
Internalizing problems				
Baseline	49	55.71 (14.09)	55.61	
6-month	41	52.6 (14.12)	52.99	2.52
12-month	25	50.62 (12.42)	50.37	
Externalizing problems				
Baseline	47	66.38 (16.43)	65.09	
6-month	42	59.59 (14.23)	61.10	4.51
12-month	28	56.88 (15.73)	57.11	
MST Pool				
Externalizing problems				
Baseline	53	72.6 (10.03)	72.28	
6-month	45	63.78 (10.93)	67.96	2.54
12-month	30	65.73 (10.39)	63.64	
Urine drug screen				
Baseline	53	21 Positive (40%)	−1.70	
6-month	36	14 Positive (25%)	−1.02	0.51
12-month	21	6 Positive (11%)	−0.96	

SE = Standard error for the model estimated slope parameter.

a pooled sample of frontline FT sessions. Pervasive adoption and implementation barriers hinder efforts to disseminate manualized FT models for adolescent behavior problems (Hogue et al., 2013). It has been argued that core elements are generally more viable and sustainable than treatment manuals for disseminating evidence-based practices in the majority of care settings (Chorpita et al., 2017; Weisz et al., 2011). This is the first study to discern the structure of core family therapy techniques as delivered by community clinicians practicing in diverse usual care

contexts. Moreover, the basic anatomy of core FT delivery in community care – four modular strategies consisting of 21 specific techniques – was found intact from that observed in controlled settings.

The current study did not entirely replicate the results of the original distillation (Hogue et al., 2019). Like that previous effort, attempts to confirm all four modules simultaneously in a single-factor analysis did not succeed. This could be due to the multiply nested structure of both data sets: Study data were collected from

Table 3. Core family therapy technique total score effects on one-year client outcomes.

	Intercept		Linear Slope			Model Fit Statistics		
	B (SE)	Pseudo-z	B (SE)	Pseudo-z	χ^2	CFI	RMSEA	SRMR
			FFT Pool (N = 50)					
Internalizing Problems	3.55 (3.66)	0.97	1.25 (2.24)	0.56	5.66 (p = .23)	0.96	0.09	–
Externalizing Problems	5.53 (5.19)	1.07	−3.28 (4.98)	−0.66	3.55 (p = .17)	0.96	0.13	–
			UC-FT Pool (N = 49)					
Internalizing Problems	−2.20 (4.36)	−0.50	−0.60 (2.61)	−0.23	3.53 (p = .47)	1.00	0.00	–
Externalizing Problems	−1.63 (4.60)	−0.35	1.40 (3.35)	−0.42	–	–	–	0.05
			MST Pool (N = 53)					
Externalizing Problems	−2.94 (1.26)	−2.33	**−3.47*** (.63)**	−5.40	–	–	–	0.13
Urine Drug Screen	−0.82 (.85)	−0.97	**0.64*** (.15)**	4.13	–	–	–	–

RMSEA = Root mean square error of approximation. CFI = Comparitive fit index. SRMR = Standardized root mean square residual.
[a]Model Chi-Square not calculated due to producing a negative Chi-Square difference test, presumably due to small sample size. CFI based on model Chi-Square, and therefore also not reported. SRMR reported in lieu of the absence of the other fit statistics and should be interpreted similar to RMSEA with ≤ .08 indicating good model fit.
[b]Conventional model fit statistics for urine drug screen outcome variable were not provided due to WLSMV estimation for categorical data.
*$p \leq$ .05, **$p \leq$ .01, ***$p \leq$ .001

treatment sessions nested within therapists nested within treatment pools. However, whereas Hogue et al. compensated for this initial failure by employing Bayesian modeling – using small variance priors placed on error variances and cross-loadings (Ozechowski, 2014) – to eventually confirm all modules and elements simultaneously, these complex modeling methods were deemed beyond the scope of this study and ultimately unnecessary given the strong evidence for factor validation in the original sample. In any event, this study's success in observationally confirming each core FT module and every constituent element in a diverse front-line sample is a rigorous finding with intriguing clinical implications.

This study also detected links between averaged FT technique use and long-term client gains. We do not want to overstate the strength of these findings: Of six technique-outcome links tested, only two effects crossed threshold for statistical significance, both occurred in the same pool, and both were small-to-medium in size. Also, the pool in which links were detected contained only two therapists, which increases potential for bias compared to the other two pools containing much larger numbers of therapists. Still, technique-outcome effects are infrequently observed even for manualized models that exercise stringent quality assurance procedures to ensure high intervention dose (Collyer et al., 2019), though FT interventions for adolescents have some record of success on this score for both manualized (e.g., M. Robbins et al., 2011) and non-manualized (Henderson et al., 2019) versions. Study findings modestly burnish this research niche.

The study also generated exploratory findings with regard to between-pool differences in delivery of core FT techniques. As might be expected FFT, a manualized FT model supported by standardized quality procedures (Robbins et al., 2019), registered the highest level of core technique use. The next highest level appeared in UC-FT, a non-manualized intervention with verified fidelity to the FT approach (Hogue, Dauber et al., 2017); followed by MST, a manualized model that includes FT techniques as a base option within a multidimensional approach that also features cognitive-behavioral techniques and, especially when treating teens with substance use problems, contingency management interventions (Randall et al., 2018; Sheidow et al., 2020). Core FT technique-outcome links were found in the MST pool only, perhaps suggesting a bonus value to emphasizing relatively greater amounts of FT interventions in treatment contexts wherein FT is not the main (or only) intervention option. This aligns with one previous study showing that greater use of core FT techniques predicted better long-term adolescent outcomes even when clients attended services featuring a predominantly non-FT treatment approach (Henderson et al., 2019). In this vein, additional research on the benefits of FT interventions as integrated components of multicomponent models, or adjunctive elements for alternative treatment approaches, would greatly serve the youth behavioral services field.

Study Strengths and Limitations

This study leveraged findings from a previous study of gold-standard sessions in controlled settings and extended those findings to more representative, community treatment contexts. The study sample overall was diverse in the race/ethnicity, sex, and geographic region of both clients and therapists. These features amplify confidence in the clinical generalizability of the distilled FT elements. By the same token, combining three community pools for factor analyses created a large degree of within-sample heterogeneity in both client and therapist characteristics, and subsample sizes for various client and therapist subgroups were not large enough to support confident investigation of their corresponding variance components. Other strengths include observational methods to measure the extensiveness of FT technique use and randomly sampled sessions from three diverse FT interventions. Given the possibility that FT techniques might have appeared at different rates in different phases of treatment in one or more pools, the study intentionally sampled from multiple available treatment phases to ensure adequate observation of all techniques. However, our demarcation of treatment phase was not intended to reflect either universal or pool-specific stages of clinical progress, nor was treatment duration standardized across pools. Also, generalizability and decision analyses suggest that upwards of ten sessions per client may be required to reliably establish therapist adherence to specific treatment models (Southam-Gerow et al., 2020); session numbers in this study fell well short of that threshold, though our goal was to sample for the appearance of various discrete techniques rather than to establish fidelity to any given model.

There were other limitations. As in the original distillation study, this study did not provide a sufficiently large sample to model the effects of community pool as a nesting variable, leaving open the possibility that one of the pools exerted a disproportionately strong or weak influence on factor analysis results. In particular, the MST pool contained only two therapists, a number that also does not allow for substantive generalization about MST model delivery. Weaknesses evident in the psychometric properties of the Relational Reframe module suggest a need to improve or reformulate the construction of this module in future studies. It was beyond study resources to assess therapist competence in delivering core FT techniques,

given that three different competency metrics would be needed to assess the degree to which therapists in each pool delivered FT in a manner both faithful to their respective treatments and responsive to individual client behaviors and needs (Kramer & Stiles, 2015). Despite sampling multiple treatment phases, analyses did not account for whether or how the FT modules or techniques were sequenced, though this "treatment coordination principle" (Chorpita et al., 2005) is an essential facet of most FT treatment manuals. Among the three FT interventions sampled in this study, only the FFT model stipulated a standardized sequence of treatment techniques. Also, the technique-outcome analyses did not attempt to discern between-phase or within-client change in technique use over the duration of therapy (e.g., Lange et al., 2019). Also, whereas it is common to operationalize adolescent substance use data as a dichotomous (yes/no) outcome variable (see Hogue et al., 2018), doing so eliminates the capacity to capture gradients of substance use severity (e.g., types of substances used, amount/frequency of use), a notable limitation given that this variable was one of only two outcomes to show significant technique-outcome links.

Clinical Implications and Future Research

This study demonstrates that the core elements of evidence-based FT manuals for adolescent conduct and substance use problems generalize to various community settings that emphasize the FT approach. These results bolster ongoing efforts to distill and disseminate core elements of family-based treatments for a variety of clinical disorders (e.g., Van Der Pol et al., 2019) and arguably lend some credibility to the idea that research-proven FT interventions can be delivered with fidelity and effectiveness in naturalistic conditions (Riedinger et al., 2017). To be sure, the community therapists in this study collectively had considerable FT experience and (for many) support from FT model experts in delivering FT interventions. Future research on training community clinicians with more limited, or no, FT experience to adopt and deliver core FT techniques will yield valuable additional data on the feasibility and potency of FT in routine care. Future studies could also focus more granularly on the differential benefits of implementing specific FT modules – that is, separately examining engagement, reframing, or interactional techniques delivered as a unified subset – in various clinical contexts (e.g., Moran et al., 2005).

Future research is also needed to test the core elements approach as a unified intervention strategy for family-based treatment of adolescent behavioral health problems, similar to the way it has been tested as a unified strategy for cognitive-behavioral therapy of childhood disorders (e.g., Chorpita et al., 2017; Weisz et al., 2012). The current study did not include any treatment model that specifically followed a unified core elements strategy; moreover, study analyses did not confirm that the four independent modules operated as a unified intervention set in the community sample pools. Core elements have also been hypothesized to exist more broadly in family and couple therapies across the lifespan (e.g., Sprenkle et al., 2013); this assertion merits empirical follow-through. If future studies on delivering core FT techniques in community settings prove promising, this would begin to mount a research-supported bid to add systemic FT interventions to the consensus roster of behavioral techniques included in training and technical support packages that aim to disseminate evidence-based practices across the youth behavioral care spectrum (e.g., Southam-Gerow et al., 2014).

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

Aaron Hogue http://orcid.org/0000-0001-8365-9545

References

Achenbach, T. M., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms & profiles: An integrated system of multi-informant assessment*. Aseba.

Asparouhov, T. (2005). Sampling weights in latent variable modeling. *Structural Equation Modeling*, 12(3), 411–434. http://dx.doi.org/10.1207/s15328007sem1203_4

Bobek, M., Porter, N., Krohner, N., & Hogue, A. (2018). *Scoring manual for core elements of family therapy: Therapist behavior rating scale (CEFT-TBRS)* Self-published document available from the authors. ahogue@toendaddiction.org

Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J. Long (Eds.), *Testing structural models*. Sage Publications.

Chorpita, B. F., Becker, K. D., & Daleiden, E. L. (2007). Understanding the common elements of evidence-based practice: Misconceptions and clinical examples. *Journal of the American Academy of Child and Adolescent Psychiatry*, 46(5), 647–652. https://doi.org/10.1097/chi.0b013e318033ff71

Chorpita, B. F., Daleiden, E., Park, A., Ward, A., Levy, M., Cromley, T., . . . Krull, J. (2017). Child STEPs in California: A cluster randomized effectiveness trial comparing modular treatment with community implemented treatment for youth with anxiety, depression, conduct problems, or traumatic stress. *Journal of Consulting and Clinical Psychology*, 85(1), 13–25. http://dx.doi.org/10.1037/ccp0000133

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(3), 566–579. http://dx.doi.org/10.1037/a0014565

Chorpita, B. F., Daleiden, E. L., & Weisz, J. R. (2005). Modularity in the design and application of therapeutic interventions. *Applied and Preventive Psychology*, 11(3), 141–156. http://dx.doi.org/10.1016/j.appsy.2005.05.002

Chorpita, B. F., Weisz, J. R., Daleiden, E. L., Schoenwald, S. K., Palinkas, L. A., Miranda, J., . . . Gibbons, R. D. (2013). Long-term outcomes for the Child STEPs randomized effectiveness trial: A comparison of modular and standard treatment designs with usual care. *Journal of Consulting and Clinical Psychology*, 81(6), 999–1009. http://dx.doi.org/10.1037/a0034200

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Laurence Erlbaum Associates.

Collyer, H., Eisler, I., & Woolgar, M. (2019). Systematic literature review and meta-analysis of the relationship between adherence, competence and outcome in psychotherapy for children and adolescents. *European Child and Adolescent Psychiatry*, 29(4), 417–431. http://dx.doi.org/10.1007/s00787-018-1265-2

Deighton, J., Argent, R., De Francesco, D., Edbrooke-Childs, J., Jacob, J., Fleming, I., . . . Wolpert, M. (2016). Associations between evidence-based practice and mental health outcomes in child and adolescent mental health services. *Clinical Child Psychology and Psychiatry*, 21(2), 287–296. https://doi.org/10.1177/1359104515589637

Diggle, P., Heagerty, P., Liang, K., & Zeger, S. (2002). *Analysis of longitudinal data* (2nd ed.). Oxford University Press.

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(4), 335. http://dx.doi.org/10.1037/ccp0000183

Duncan, T. E., Duncan, S. C., Strycker, A. L., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling*. Sage.

Fleiss, J. (1981). Balanced incomplete block designs for inter-rater reliability studies. *Applied Psychological Measurement*, 5(1), 105–112. https://doi.org/10.1177/014662168100500115

Gillespie, M. L., Huey, S. J., & Cunningham, P. B. (2017). Predictive validity of an observer-rated adherence protocol for multisystemic therapy with juvenile drug offenders. *Journal of Substance Abuse Treatment*, 76, 1–10. http://dx.doi.org/10.1016/j.jsat.2017.01.001

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(3), 308. http://dx.doi.org/10.1037/ccp0000376

Henggeler, S. W., & Schaeffer, C. M. (2016). Multisystemic therapy®: Clinical overview, outcomes, and implementation research. *Family Process*, 55(3), 514–528. http://dx.doi.org/10.1111/famp.12232

Hogue, A., Bobek, M., Dauber, S., Henderson, C. E., McLeod, B. D., & Southam-Gerow, M. A. (2017). Distilling the core elements of family therapy for adolescent substance use: Conceptual and empirical solutions. *Journal of Child & Adolescent Substance Abuse*, 26(6), 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(1), 29–41. https://doi.org/10.1080/15374416.2018.1555762

Hogue, A., Dauber, S., Henderson, C., Bobek, M., Johnson, C., Lichvar, E., & Morgenstern, J. (2015). Randomized trial of family therapy versus non-family treatment for adolescent behavior problems in usual care. *Journal of Clinical Child and Adolescent Psychology*, 44(6), 954–969. https://doi.org/10.1080/15374416.2014.963857

Hogue, A., Dauber, S., & Henderson, C. E. (2017). 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(5), 626–641. http://dx.doi.org/10.1007/s10488-016-0769-7

Hogue, A., Dauber, S., Samuolis, J., & Liddle, H. (2006). Treatment techniques and outcomes in multidimensional family therapy for adolescent behavior problems. *Journal of Family Psychology*, 20(4), 535–543. https://doi.org/10.1037/0893-3200.20.4.535

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(4), 499–526. https://doi.org/10.1080/15374416.2018.1466307

Hogue, A., Henderson, C. E., Dauber, S., Barajas, P., Fried, A., & Liddle, H. (2008). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology*, 76(4), 544–555. https://doi.org/10.1037/0022-006X.76.4.544

Hogue, A., & Liddle, H. (2009). Family-based treatment for adolescent substance abuse: Controlled trials and new horizons in services research. *Journal of Family Therapy*, 31(2), 126–154. http://dx.doi.org/10.1111/j.1467-6427.2009.00459.x

Hogue, A., Ozechowski, T., Robbins, M., & Waldron, H. (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(1), 60–77. http://dx.doi.org/10.1111/cpsp.12023

Kline, P. (1979). *Psychometrics and Psychology*. Academic 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(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kramer, U., & Stiles, W. B. (2015). The responsiveness problem in psychotherapy: A review of proposed solutions. *Clinical Psychology: Science and Practice*, 22(3), 277–295. http://dx.doi.org/10.1111/cpsp.12107

Lange, A. M., van der Rijken, R. E., Delsing, M. J., Busschbach, J. J., & Scholte, R. H. (2019). Development of therapist adherence in relation to treatment outcomes of adolescents with behavioral problems. *Journal of Clinical Child and Adolescent Psychology*, 48(sup1), S337–S346. https://doi.org/10.1080/15374416.2018.1477049

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

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936–949. http://dx.doi.org/10.3758/s13428-015-0619-7

Liddle, H. A. (2016). Multidimensional family therapy: Evidence base for transdiagnostic treatment outcomes, change mechanisms, and implementation in community settings. *Family Process*, 55(3), 558–576. http://dx.doi.org/10.1111/famp.12243

Little, R., & Rubin, D. (2002). *Statistical analysis with missing data* (2nd ed.). Wiley. http://doi.org/10.1002/9781119013563

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(5), 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(2), 180–200. http://dx.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(5), 529–563. https://doi.org/10.1080/15374416.2016.1146990

McDonald, R., & Ho, M. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, 7(1), 64–82. https://doi.org/10.1037/1082-989X.7.1.64

McHugh, R., & Barlow, D. (2010). The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist*, 65(2), 73–84. http://dx.doi.org/10.1037/a0018121

McLeod, B. D., Jensen-Doss, A., & Ollendick, T. H. (2013). Overview of diagnostic and behavioral assessment. In B. D. McLeod, A. Jensen-Doss, & T. H. Ollendick (Eds.), *Diagnostic and behavioral assessment in children and adolescents: A clinical guide* (pp. 3–33). Guilford Publications, Inc.

Minuchin, S., & Fishman, H. C. (1981). *Family therapy techniques*. Harvard.

Moran, G., Diamond, G. M., & Diamond, G. S. (2005). The relational reframe and parents' problem constructions in attachment-based family therapy. *Psychotherapy Research*, 15(3), 226–235. https://doi.org/10.1080/10503300512331387780

Muthén, L. K., & Muthén, B. O. (1998/2017). *Mplus user's guide* (7th ed.). Muthen & Muthen.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw Hill.

Ozechowski, T. J. (2014). Empirical Bayes MCMC estimation for modeling treatment processes, mechanisms of change, and clinical outcomes in small samples. *Journal of Consulting and Clinical Psychology*, 82(5), 854. http://dx.doi.org/10.1037/a0035889

Randall, J., Cunningham, P. B., & Henggeler, S. W. (2018). The development and transportability of multisystemic therapy-substance abuse: A treatment for adolescents with substance use disorders. *Journal of Child & Adolescent Substance Abuse*, 27(2), 59–66. https://doi.org/10.1080/1067828X.2017.1411301

Riedinger, V., Pinquart, M., & Teubert, D. (2017). Effects of systemic therapy on mental health of children and adolescents: A meta-analysis. *Journal of Clinical Child and Adolescent Psychology*, 46(6), 880–894. https://doi.org/10.1080/15374416.2015.1063427

Robbins, M., Feaster, D., Horigian, V., Puccinelli, M., Henderson, C. E., & Szapocznik, J. (2011). Therapist adherence in brief strategic family therapy for adolescent drug abusers. *Journal of Consulting and Clinical Psychology*, 79(1), 43–53. http://dx.doi.org/10.1037/a0022146

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(3), 543–557. http://dx.doi.org/10.1111/famp.12230

Robbins, M. S., Waldron, H. B., Turner, C. W., Brody, J., Hops, H., & Ozechowski, T. (2019). Evaluating supervision models in functional family therapy: Does adding observation enhance outcomes? *Family Process*, 58(4), 873–890. http://dx.doi.org/10.1111/famp.12399

Sheidow, A. J., Zajac, K., Chapman, J. E., McCart, M. R., & Drazdowski, T. K. (2020). Randomized controlled trial of an integrated family-based treatment for adolescents presenting to community mental health centers. *Community Mental Health Journal*. https://doi.org/10.1007/s10597-020-00735-z

Shrout, P., & Fleiss, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Southam-Gerow, M. A., Bonifay, W., McLeod, B. D., Cox, J. R., Violante, S., Kendall, P. C., & Weisz, J. R. (2020). Generalizability and decision studies of a treatment adherence instrument. *Assessment*, 27(2), 321–333. https://doi.org/10.1177/1073191118765365

Southam-Gerow, M. A., Daleiden, E. L., Chorpita, B. F., Bae, C., Mitchell, C., Faye, M., & Alba, M. (2014). MAPping Los Angeles County: Taking an evidence-informed model of mental health care to scale. *Journal of Clinical Child and Adolescent Psychology*, 43(2), 190–200. https://doi.org/10.1080/15374416.2013.833098

Sprenkle, D. H., Davis, S. D., Lebow, J. L., & Lebow, J. (2013). *Common factors in couple and family therapy: The overlooked foundation for effective practice*. Guilford.

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(2), 145–158. http://dx.doi.org/10.1016/j.jsat.2012.05.006

Van Der Pol, T. M., van Domburgh, L., Van Widenfelt, B. M., Hurlburt, M. S., Garland, A. F., & Vermeiren, R. R. (2019). Common elements of evidence-based systemic treatments for adolescents with disruptive behaviour problems. *The Lancet Psychiatry*, 6(10), 862–868. http://dx.doi.org/10.1016/S2215-0366(19)30085-9

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(2), 200–211. http://dx.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(1), 11–43. https://doi.org/10.1080/15374416.2016.1220310

Weisz, J. R., Bearman, S., Santucci, L., & Jensen-Doss, A. (2017). Initial test of a principle-guided approach to trans-diagnostic psychotherapy with children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, 46(1), 44–58. https://doi.org/10.1080/15374416.2016.1163708

Weisz, J. R., Bearman, S. K., Ugueto, A. M., Herren, J. A., Evans, S. C., Cheron, D. M., . . . Jensen-Doss, A. (2020). Testing robustness of child STEPs effects with children and adolescents: A randomized controlled effectiveness trial. *Journal of Clinical Child & Adolescent Psychology*, 49(6), 883–896. https://doi.org/10.1080/15374416.2019.1655757

Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S. K., Miranda, J., Bearman, S. K., . . . Gibbons, R. D. (2012). Testing standard and modular designs for psychotherapy treating depression, anxiety, and conduct problems in youth: A randomized effectiveness trial. *Archives of General Psychiatry*, 69(3), 274–282.

Weisz, J. R., Ugueto, A. M., Herren, J., Afienko, S. R., & Rutt, C. (2011). Kernels vs. ears, and other questions for a science of treatment dissemination. *Clinical Psychology: Science and Practice*, 18, 41–46. http://dx.doi.org/10.1111/j.1468-2850.2010.01233.x

Weisz, J. R., Ugueto, A. M., Herren, J., Marchette, L. K., Bearman, S. K., Lee, E. H., Alleyne, A., Cheron, D. M., Tweed, J. L., Hersh, J., Raftery-Helmer, J. N., Weissman, A. S., Jensen-Doss, A., & Thomassin, K. (2018). When the torch is passed, does the flame still burn? Testing a "train the supervisor" model for the Child STEPs treatment program. *Journal of Consulting and Clinical Psychology*, 86(9), 726. http://dx.doi.org/10.1037/ccp0000331