Title

Can a Single Measure Estimate Protocol Adherence for Two Psychosocial Treatments for Youth Anxiety Delivered in Community Mental Health Settings?

Journal

Behavior Therapy, 53(1)

Authors

McLeod, Bryce
Martinez, Ruben
Southam-Gerow, Michael
et al.

Publication Date

2022

DOI

10.1016/j.beth.2021.06.008

Peer reviewed

HHS Public Access

Author manuscript

Behav Ther. Author manuscript; available in PMC 2023 January 01.

Published in final edited form as:

Behav Ther. 2022 January ; 53(1): 119–136. doi:10.1016/j.beth.2021.06.008.

Can a Single Measure Estimate Protocol Adherence for Two Psychosocial Treatments for Youth Anxiety Delivered in Community Mental Health Settings?

Bryce D. McLeod,
Department of Psychology, Virginia Commonwealth University.

Ruben G. Martinez,
University of California, Los Angeles

Michael A. Southam-Gerow,
Department of Psychology, Virginia Commonwealth University.

John R. Weisz,
Harvard University

Bruce F. Chorpita
University of California, Los Angeles.

Abstract

Treatment adherence measurement can be time and resource-intensive in clinical trials, so the ability to measure protocol adherence for two distinct treatment programs with a single measure may benefit the field. The present study sought to determine if the Therapy Process Observational Coding System – Revised Strategies Scale (TPOCS-RS) could assess protocol adherence to two youth treatment programs. Treatment sessions (N = 796) from 55 youth (M age = 9.89 years, SD = 1.71; range 7–15 years; 55.0% White; 46.0% female) with primary anxiety problems treatment by 39 clinicians (M age = 40.54 years, SD = 9.56; 50.0% White; 80.0% female) were independently scored by coders using observational treatment adherence and alliance measures. The youth received one of three treatments: (a) Standard (i.e., cognitive-behavioral treatment program), (b) Modular (i.e., a program with cognitive-behavioral and parent training components), or (c) Usual Care. Consultants filled out a self-report measure of protocol adherence within the Standard and Modular conditions. Interrater reliability, ICC(2,2) for the various items for the full sample ranged from .17 to .92 (M ICC = .67; SD = .17). Scores from a TPOCS-RS subscale that mapped onto the specific content of the treatment protocols used in the Standard and Modular conditions evidenced convergent validity with the consultant-report adherence measure and discriminant validity with the alliance measure. The model-specific TPOCS-RS subscales also discriminated between the Standard and Modular treatments and Usual Care. This study provides initial evidence that (a) the

Correspondence concerning this article should be addressed to Bryce D. McLeod, Department of Psychology, Virginia Commonwealth University, 806 West Franklin Street, PO Box 842018, Richmond, VA 23284- 2018. bmcleod@vcu.edu.

TPOCS-RS has utility in estimating protocol adherence in different treatment programs and (b) support the score validity of the self-report consultation records.

Keywords

treatment integrity; cognitive-behavioral therapy; youth anxiety

Psychosocial treatment evaluation and implementation research both emphasize the importance of assessing treatment integrity (Cox et al., 2019; Proctor et al., 2011). Treatment integrity definitions often include three components focused on the delivery of therapeutic techniques (e.g., a discrete practice delivered by a clinician such as exposure) that are specified in treatment protocols (Regan et al., 2019). These components include *adherence* (i.e., delivery of techniques specified in a protocol), *differentiation* (i.e., delivery of techniques not found in the protocol), and *competence* (i.e., how well the techniques specified in a protocol are delivered; Schoenwald et al., 2011). Though the measurement of treatment integrity is important, the methods and measures used to assess treatment integrity vary greatly (Cox et al., 2019).

Most treatment integrity measures are designed to assess adherence to a protocol for a specific youth emotional or behavioral problem (Cox et al., 2019). Typically, these measures were developed to facilitate manipulation checks (i.e., a test to ascertain if a variable was successfully manipulated) intended to determine if the techniques from a specific psychosocial treatment (hereafter called *treatment*) was delivered as quantified in the protocol (Cox et al., 2019). This is called *protocol adherence* and is defined as determining the extent to which the techniques specified in a protocol are delivered by a clinician over treatment. Determining protocol adherence can be useful early in treatment development or during efficacy testing (Carroll & Nuro, 2002). However, measures designed to capture specific techniques from a particular protocol may not be a good fit for all research questions (Schoenwald et al., 2011).

Treatment-treatment comparisons are commonplace in effectiveness and implementation research (see Weisz et al., 2012). Such studies may include comparisons between two treatments (e.g., Haugland et al., 2020), a treatment and usual care (e.g., Southam-Gerow et al., 2010), or some combination thereof (e.g., Weisz et al., 2012). Measurement of protocol adherence in such studies can serve as a manipulation check or as a dependent variable (i.e., if training and consultation were successful; Proctor et al., 2011).

Beyond protocol adherence, measurement of differentiation can help answer certain questions relevant to effectiveness and implementation research. When clinicians in community settings are trained to deliver a specific treatment program, they may deliver techniques that are not included or are proscribed (i.e., expressly forbidden) by the treatment protocol (e.g., the use of psychoanalytic techniques in cognitive-behavioral therapy [CBT]; Smith et al., 2017). Moreover, usual care, defined as the treatment typically provided and believed to be effective by clinicians in community-based mental health settings, can contain a wide range of techniques that are associated with various theoretical orientations (Brookman-Frazee et al., 2010; Garland et al., 2010), including techniques found in

evidence-based treatments (Smith et al., 2017). As it is difficult to predict what techniques may be delivered by clinicians in community settings, assessing differentiation can aid interpretation of findings by allowing researchers to gauge “treatment purity” (Schoenwald et al., 2011) or characterize treatment modifications (i.e., changes made to a treatment to promote fit with a new service context; Marques et al., 2019).

Though the assessment of protocol adherence and differentiation helps to answer certain research questions, the design of existing measures can make the efficient measurement of protocol adherence and differentiation difficult (Schoenwald et al., 2011). Most treatment integrity measures are designed to gauge protocol adherence for a specific treatment (Southam-Gerow et al., 2016), so more than one measure may be needed to assess protocol adherence and differentiation (Collyer et al, 2019; Cox et al., 2019). For example, a treatment-treatment comparison may require the use of two treatment-specific adherence measures and a differentiation measure. A single measure that assesses both protocol adherence and differentiation may thus offer some benefits.

Recognizing the need for more efficient measurement options, some researchers have moved away from developing treatment integrity measures for a specific treatment model (e.g., Garland et al., 2010; McLeod et al., 2015; Southam-Gerow et al., 2016). Instead of including items developed based on the content of a particular treatment protocol, items assess discrete techniques found across multiple treatments for one or more youth emotional and behavioral problems (e.g., Garland, et al., 2010; Southam-Gerow et al., 2016). A measure designed to capture a wide range of techniques from multiple treatment approaches (e.g., CBT, family therapy, parent-training) for various problem types may allow researchers to use a single measure to assess both protocol adherence and differentiation.

The Therapy Process Observational Coding System-Revised Strategy scale (TPOCS-RS; McLeod et al., 2015) has design features that may allow researchers to use it to assess protocol adherence and differentiation. The measure is designed to provide extensiveness estimates, based on a 7-point Likert-type scale (Hogue et al., 1996), for techniques from five theory-based domains (e.g., cognitive, psychodynamic). These ratings provide quantity, or dosage, information about each technique, which is appropriate for estimating protocol adherence and differentiation, but do not provide estimates of competence. Also, the TPOCS-RS is not problem specific, so it can be used to characterize treatment delivery for different youth behavioral and emotional problems (e.g., Southam-Gerow et al., 2010). Originally designed to characterize the techniques delivered in usual care, the TPOCS-RS has shown promise as a differentiation measure (McLeod et al., 2015; Smith et al., 2017). Though the TPOCS-RS items were not designed to assess techniques found in specific treatment protocols, it may be possible to estimate protocol adherence by mapping TPOCS-RS items onto the content of specific treatment protocols. If the TPOCS-RS shows potential in assessing protocol adherence to specific treatment protocols, then for certain research questions it could provide a way to assess both protocol adherence and differentiation with a single measure.

The goal of this study was to evaluate the potential of the TPOCS-RS to assess protocol adherence for two distinct treatment programs delivered in community settings. To estimate

protocol adherence, the TPOCS-RS items and subscale scores need to demonstrate several reliability and validity features. First, TPOCS-RS items that map onto the content of a treatment protocol would need to demonstrate adequate interrater reliability (ICC > .40; Cicchetti, 1994). Second, a TPOCS-RS subscale comprised of items that map onto the content of each treatment protocol would need to evidence overlap with scores on measures designed to gauge protocol adherence for a specific treatment program (i.e., convergent validity). Third, variation in scores on the TPOCS-RS subscale comprised of items that map onto the content of a protocol need to replicate the pattern of associations with other TPOCS-RS subscales and the alliance reported in previous integrity research (i.e., discriminant validity; e.g., McLeod et al., 2015). Finally, the TPOCS-RS subscales comprised of items that map onto the content of each treatment protocol would need to detect expected differences between each program and usual care (i.e., discriminative validity; Weisz et al., 2012).

To achieve study goals, the TPOCS-RS was used to code treatment sessions for youth with primary anxiety problems who participated in a randomized effectiveness trial. Youth and clinicians were randomized to a standard manual condition who delivered the Coping Cat program (Kendall & Hedtke, 2006), to a modular condition that delivered the Modular Approach to Therapy with Children (MATCH; Chorpita & Weisz, 2005), or to usual care. Coping Cat is a child-focused CBT program designed for youth with anxiety disorders whereas MATCH is a set of CBT and parent-management training techniques designed for youth with anxiety, depression, and conduct problems. As youth in this study presented with primary anxiety problems, clinicians delivered MATCH used CBT techniques to address anxiety problems, though other CBT and parent management techniques could be referenced if needed during treatment to address other problems. Though previous research has examined the score reliability and validity of the TPOCS-RS (e.g., McLeod et al., 2015; McLeod & Weisz, 2010), this is the first study to examine the potential of the TPOCS-RS to estimate protocol adherence.

Method

Data Sources and Participants

Treatment data were collected from 55 youth (*M* age = 9.89 years, *SD* = 1.71; range 7–15 years; 55.0% White; 46.0% female, 54.0% male) and 39 clinicians (*M* age = 40.54 years, *SD* = 9.56; 50.0% White; 80.0% female, 20.0% male) who participated in a randomized effectiveness trial called the Child STEPS Multisite Trial (Weisz et al., 2012) conducted between 2005 and 2010. To be eligible for the Child STEPS Multisite Trial, youth had to meet DSM-IV-TR (American Psychiatric Association, 2000) criteria according to the Children's Interview for Psychiatric Syndromes (CHIPS; Weller et al., 1999a,b) or have clinically elevated problems (T-score > 65) on the Child Behavior Checklist or Youth Self Report (Achenbach & Rescorla, 2000) in at least one of three areas: anxiety, depression, or conduct problems. The Child STEPS Multisite Trial used a cluster randomization design. Clinicians were assigned to one of three conditions randomly, with stratification by educational level (master's vs doctoral degree): standard manualized treatment (Standard), modular manualized treatment (Modular), or usual care. Youth assigned to a study clinician

as part of the standard clinic referral process was able to be screened, and if found to be eligible for the study, invited to consent.

For inclusion in the current study, youth had to (a) present with a primary anxiety problem (i.e., CHIPS diagnosis of specific phobia, separation anxiety disorder, generalized anxiety disorder, social phobia, obsessive-compulsive disorder, posttraumatic stress disorder, panic disorder without agoraphobia or a T-score > 65 on the CBCL Anxious-Depressed subscale), (b) have at least two audible recorded sessions, and (c) have received treatment from a single clinician. A total of 13 youth with primary anxiety problems and six clinicians from the Child STEPS Multisite Trial did not meet these inclusion criteria. Video- and audio-recorded sessions collected in the Child STEPS Multisite Trial served as the data for the present study. The study was institutional review board approved. Parents provided written informed consent, and youth gave written or oral assent.

Out of the 55 youths with a primary anxiety problem who met inclusion criteria for the current study, 22 youths were allocated to the Standard condition (*M age* = 9.77 years; *SD* = 1.51; 50.0% female, 50.0% male; 72.7% White), 16 youths to the Modular condition (*M* age = 9.94 years; *SD* = 1.88; 43.7% female, 56.3% male; 43.8% White), and 17 youths to the Usual Care condition (*M* age = 10.00 years; *SD* = 1.87; 41.2% female, 58.8% male; 41.2% White). Findings from Child STEPS Multisite Trial indicated that at post-treatment youth in the Modular condition had better outcomes on multiple clinical measures than youth in the Standard and Usual Care conditions. See Table 1 for descriptive information.

The 39 clinicians who volunteered to participate were randomly assigned to condition. The 16 clinicians (*M* age = 43.56 years; *SD* = 9.96; 82.2% female, 18.8% male; 50.0% White) in the Standard condition averaged 7.17 years (*SD* = 7.75) of clinical experience and included 37.5% master's-level social workers, 31.3% master's-level psychologists, 6.3% doctoral-level psychologists, and 25.0% classified as "other" (e.g., marriage and family therapist). The 10 clinicians (*M* age = 35.20 years; *SD* = 6.81; 80.0% female, 20.0% male; 40.0% White) in the Modular condition averaged 5.25 years (*SD* = 4.83) of clinical experience and included 50.0% master's-level social workers, 20.0% master's-level psychologists, 10.0% doctoral-level psychologists, and 20.0% classified as "other." The 13 clinicians (*M* age = 40.00 years; *SD* = 9.18; 76.9%, 23.1% male; 61.5% White) in the Usual Care condition averaged 4.69 years (*SD* = 5.34) of clinical experience and included 38.5% master's-level social workers, 23.1% master's-level psychologists, 30.7% doctoral-level psychologists, and 7.7% classified as "other." See Table 2 for clinician descriptive information.

Treatment Conditions

Standard.—The Coping Cat program, a child-focused CBT program for youth diagnosed with anxiety disorders (Kendall & Hedtke, 2006), was delivered in the Standard condition. Coping Cat consists of 16 sessions; 14 sessions are conducted individually with the youth, and two sessions are conducted with the caregiver. The first half of treatment focuses on anxiety management skills training (e.g., problem solving), whereas the second half emphasizes exposures. Sessions are designed to be delivered in a predetermined order for each youth. If at the end of the Coping Cat program a secondary depressive or disruptive behavior disorder was present, a second treatment for depression (i.e., Primary

and Secondary Control Enhancement Training [PASCET]) or disruptive behavior (i.e., Defiant Children) was delivered.

Modular.—MATCH (Chorpita & Weisz, 2005) was delivered in the Modular condition. MATCH is comprised of modules designed to address anxiety, depression, and conduct problems. The modules are comprised of CBT and parent-training techniques that correspond to those found in (a) Coping Cat (Kendall & Hedtke, 2006), (b) PASCET, a CBT program for depression (Weisz et al., 1999), and (c) Defiant Children, a behavioral parent training program for conduct problems (Barkley, 2013). Flowcharts specify a default order of modules for each problem area. Youth and caregiver scores on baseline measures were used to identify the primary problem area and the corresponding flowchart. As the current study focused on youth with primary anxiety problems, all of the clinicians used the flowchart associated with anxiety. However, if a crisis, stressor, or comorbid condition arose during treatment, clinicians were free to reference the flowcharts to incorporate modules designed to address those conditions, permitting a return to the focus on anxiety. The treatment protocol likewise enabled a clinician to shift the treatment focus, if evidence indicated that another problem warranted primary consideration (e.g., depression).

Usual care.—Clinicians in the Usual Care condition continued their normal procedure for conducting treatment, including the frequency and model of supervision (see Weisz et al., 2012). The clinicians in Usual Care reported the following theoretical orientations: 30.8% CBT, 30.8% eclectic, 15.4% psychodynamic, 15.4% “other,” and 7.7% family systems.

Training and consultation.—The same training and consultation procedures were used in the Standard and Modular conditions and included a treatment protocol, training workshop, and weekly consultation with an expert. Training was held with all clinicians over six days, with two days allocated to each problem area (anxiety, depression, and conduct). Postdoctoral project consultants, all with PhDs in clinical psychology and all with training and clinical experience in CBT, were trained in MATCH components by experts in each treatment protocol. The consultants engaged the clinicians in weekly discussions that included review of measurement feedback on client progress and practices delivered (Chorpita et al., 2008). A treatment protocol adherence check revealed that 92.8% of session content in the Standard condition was model specific and 7.2% of content was not part of the protocol; 82.9% of session content in the Modular condition was specific to the protocol and 17.1% of content was not part of the protocol. In Usual Care, 91.4% of content was not found in the protocols used in the Standard or Modular conditions, indicating that the majority of the techniques delivered in Usual Care were distinct from those delivered in the Standard and Modular conditions. See Weisz et al. (2012) for details regarding the training, consultation, and protocol adherence checks.

Instrumentation

Therapy Process Observational Coding System for Child Psychotherapy—Revised Strategies scale—(TPOCS-RS; McLeod et al., 2015) is a 47-item measure designed to assess clinician delivery of techniques across five theory-based subscales: Cognitive (4 items; e.g., “Cognitive Distortions”), Behavioral (9 items; e.g., “Operant

Strategies–Child"), Psychodynamic (5 items; e.g., "Interpretation"), Family (8 items; e.g., "Parenting Skills"), and Client-Centered (4 items; e.g., "Positive Regard"). Seventeen additional items (e.g., "Homework") represent techniques that play a meaningful role in treatment but are not associated with a subscale. Four items were added to the TPOCS-RS: "Crisis Management" (i.e., helps client address an urgent event), "Case Management" (i.e., activities designed to coordinate services designed to supplement treatment), "Therapy Engagement" (i.e., addresses barriers to treatment participation), and "Educational Support" (i.e., exercises to assist youth with academic problems). Feedback from coders who had used the TPOCS-RS (see McLeod et al., 2015) to code sessions conducted in different settings (research vs clinical) and representing various treatment types (CBT vs Usual Care) led to the addition of these four items. These items expanded the number of delivery methods and incorporated supportive techniques (e.g., "Case Management"). An item was added to the Family subscale: "Operant Strategies–Parent" (i.e., teaches principles of or encourages parents to use an operant system), which allowed for the separation of operant techniques targeted at the youth ("Operant Strategies–Child"; i.e., therapist teaches principles of operant system or uses an operant system with the youth) from those directed at the caregiver. Coders rate the extent to which the clinician engages in each technique during an entire session on a 7-point extensiveness scale: 1 = "not at all" to 7 = "extensively." The TPOCS-RS, or variants thereof (i.e., PRAC-TPOCS, Garland et al., 2010; TPOCS-S, McLeod & Weisz, 2010), have demonstrated mean item interrater reliability ranging from ICC .59 to .86 (e.g., McLeod et al., 2015; McLeod & Weisz, 2010), provided evidence of construct validity in research- and community-based samples (McLeod et al., 2015; Southam-Gerow et al., 2016), differentiated between treatment types (e.g., Southam-Gerow et al., 2010), and demonstrated predictive validity (Garland et al., 2014). See Table 3 for a definition of each TPOCS-RS item; the TPOCS-RS scoring manual is available from the first author.

The TPOCS-RS "Anxiety" subscale was used to estimate protocol adherence in each condition, since the techniques used to address primary anxiety problems were the same in the Standard and Modular conditions. This subscale, used in previous studies focused on the Coping Cat program (see McLeod et al., 2015; Smith et al., 2017; Southam-Gerow et al., 2010), contains six cognitive and behavioral items that map onto the content of CBT programs for youth anxiety (see Table 3): "Cognitive Education," "Cognitive Distortion," "Coping Skills," "Relaxation," "Operant-Child," and "Respondent." To facilitate the evaluation of discriminant validity, we created additional condition-specific subscales. First, we generated the TPOCS-RS Non-Standard subscale comprised of the child-focused CBT techniques not used in the Anxiety subscale: "Functional Analysis," "Skill Building," and "Behavioral Activation." Second, for the Modular condition we created two subscales (a) a TPOCS-RS Non-Modular subscale comprised of the "Functional Analysis" and "Skill Building" items—"Behavioral Activation" was not included because it is included in the MATCH protocol as part of techniques used to address depression; and (b) a modified TPOCS-RS Family subscale used for analyses involving the Modular condition comprised of "Recruits Others," "Parenting Style," "Multiparticipant Interactions," and "Family Member Roles"—the "Targets Others," "Operant Strategies-Parent," and "Parenting Skills" items were not included because these techniques are in the MATCH protocol to address conduct problems. Scores on each TPOCS-RS subscale, including the Psychodynamic,

Family, Client-Centered subscales, were generated by first producing a mean score on each item across coders and then averaging the item scores on each subscale.

MATCH and Standard Consultation Records—(Ward et al., 2013) represent two consultant-report measures designed to capture what techniques were delivered in a session in the Standard and Modular conditions within the Child STEPs Multisite Trial (Ward et al., 2013). Items on the Standard Consultation Record (i.e., Consultation Standard) consist of techniques of the Coping Cat protocol (e.g., "FEAR Plan,"), while items on the MATCH Consultation Record (i.e., Consultation Modular Anxiety) consist of the components of the MATCH protocol (e.g., "Cognitive STOP,"). Data were gathered during weekly meetings between a clinician and project consultants via a collaborative semi-structured interview where a clinician reported the practices used and the consultant asked follow-up questions. Scores on the Consultation Standard and Consultation Modular Anxiety scales have demonstrated good interrater reliability (ICCs ranged from .50 to 1.00; M = .80) when rated by independent observers, and evidence of convergent validity when compared to consultation records scored by independent observers (Ward et al., 2013). For each item, the consultant could select "no selection," "covered-part" (partial coverage of the session content), or "covered-full" (full coverage of the session content). For this study, scores were recoded into a 3-point scale such that no selection = 0, "covered-part" = 1, and "covered full" = 2. For the Standard condition, we used the Consultation Standard scale comprised of the items focused on anxiety techniques. For the Modular condition, we created a Consultation Modular Anxiety scale comprised of the items focused on techniques for anxiety. Scale scores for each session were created by averaging the items on each scale.

Therapy Process Observational Coding System for Child Psychotherapy—Alliance scale—(TPOCS-A; McLeod & Weisz, 2005) assesses the quality of the client-clinician alliance in youth treatment. Six items assess affective elements of the relationship, and three items assess client participation in therapeutic activities. Coders observe entire sessions and rate each item on a 6-point scale ranging from 0 ("not at all") to 5 ("a great deal"). The TPOCS-A has demonstrated item interrater reliability ranging from .48 to .80 (M ICC = .67) and internal consistency ranging from .91 to .95 (M α = .92). Scores on the TPOCS-A scale have demonstrated evidence of convergent validity with self-report alliance measures ranging from rs .48 to .53 (e.g., Fjermestad et al., 2012; Liber et al., 2010) and predictive validity with clinical outcomes (Liber et al., 2010; McLeod & Weisz, 2005). Interrater reliability, ICC(2,2), for the TPOCS-A in the present sample was .85; internal consistency was .89.

Coding and Session Sampling Procedures

Two doctoral students in clinical psychology comprised the TPOCS-RS coding team (100.0% female; M age 27.00 years, SD = 1.41; 50.0% White, 50.0% Latina), and three doctoral students in clinical psychology (M age = 26.00 years, SD = 2.00; 100.0% female; 66.7% White, 33.3% Asian-American) comprised the TPOCS-A team. Training progressed through the same steps for each measure. First, coders received detailed instruction in the scoring manual, reviewed sessions with the trainers, and engaged in coding exercises designed to expand understanding of each item. Second, coders engaged in both group

and independent coding and discussed results in weekly meetings. Lastly, each coder independently coded 40 sessions; reliability was assessed against master codes. To be certified for independent coding, each coder had to demonstrate "good" reliability on each item (ICC > .59; Cicchetti, 1994). Once trained, coders typically scored a treatment session in about 47 minutes (*M* session length = 41.39 minutes, *SD* = 12.58 + 5 minutes; see Table 1). All available sessions for each case were coded except the first and last session as these sessions may contain intake or termination content. A total of 876 sessions were held; 796 (90.8%) were rated (94.7% Standard, *n* = 359; 93.8% Modular, *n* = 244; 81.4% Usual Care, *n* = 193). Sessions were not rated if: (a) less than 15 minutes were audible, or (b) the recording was missing. Coding order was determined by random assignment. Each session was double-coded. Coders were naïve to study hypotheses.

Assessments Collected in the Original RCT

The *Child Behavior Checklist* (CBCL; Achenbach & Rescorla, 2000) was used to assess symptomatology in the Child STEPS Multisite Trial. In the current investigation, t-scores on three CBCL scales were used for descriptive purposes and group comparisons: Total, Internalizing (broadband), Externalizing (broadband).

Results

Preliminary Analyses

We conducted sample bias analyses to ascertain if the 55 youths and 39 clinicians selected for the current study differed from the other participants in the anxiety subsample of Child STEPs Multisite Trial (see Weisz et al., 2012). Thirteen youth participants with primary anxiety problems and six clinicians from the Child STEPS Multisite Trial were not included in the current study because they did not meet inclusion criteria. The excluded youths and clinicians did not differ from youth and clinicians in the current sample. For the sample of 55 youths, comparisons across the Standard, Modular, and Usual Care conditions for youth and clinician characteristics were examined. Only one significant difference was found, sessions in the Modular condition (*M* = 43.34 minutes, *SD* = 12.27) were significantly longer than the sessions in the Standard (*M* = 40.49 minutes, *SD* = 13.54; t = 2.63, *p* = .009) and Usual Care conditions (*M* = 40.65 minutes, *SD* = 10.70; t = 2.40, *p* = .02). Finally, we examined whether the same percentage of sessions was coded across the three conditions. There was not a significant difference in percent of sessions coded between the three conditions, *F*(1, 2) = .22, *p* = .80. See Tables 1 and 2 for sample comparisons.

Interrater Reliability

We investigated the interrater reliability for each TPOCS-RS item. Interrater reliability was estimated using ICC(2,2), which is based on a two-way random effects model (Shrout & Fleiss, 1979). ICC(2,2) provides a reliability estimate of the mean scores of coders considered as a whole and allows for generalizability of the results. Interrater reliability was estimated for the full sample and then separately for each condition. Following Cicchetti (1994), ICCs values below .40 reflect "poor" agreement, ICCs from .40 to .59 reflect "fair" agreement, ICCs from .60 to .74 reflect "good" agreement and ICCs .75 and higher reflect "excellent" agreement.

ICCs for the various items, based on the full sample, ranged from .17 to .92 (*M* ICC = .67; *SD* = .17). Of the 47 items, 16 were in the "excellent" range, 13 in the "good" range, 14 in the "fair" range, and 4 in the "poor" range ("Explores Past," "Recruits Others," "Advice," "Therapy Engagement"). Four of the five new items evidenced adequate interrater reliability; "Therapy Engagement" was the one exception with an ICC below .40. These findings indicate that 43 of the 47 TPOCS-RS items in the full sample demonstrated fair to excellent interrater reliability.

As seen in Table 3, interrater reliability of the items varied in each condition. In the Standard condition, the interrater reliability for the TPOCS-RS items ranged from −.01 to .93 (*M* ICC = .60, *SD* = .23), with five items falling in the "poor" range (Cicchetti, 1994). The interrater reliability for the TPOCS-RS items in the Modular condition ranged from .00 to .94 (*M* ICC = .60, *SD* = .23), with five items in the "poor" range. Within the Usual Care condition, the interrater reliability for the TPOCS-RS items ranged from −.02 to .87 (*M* ICC = .57, *SD* = .23), with 10 items falling within the "poor" range. Items on the TPOCS-RS Anxiety subscale evidenced fair to excellent interrater reliability in the Standard and Modular conditions, though two of these items evidence poor reliability in Usual Care (i.e., "Coping Skills," "Respondent Interventions"). Overall, most TPOCS-RS items evidenced adequate interrater reliability in the three conditions.

Based on the interrater reliability generated for the full sample, two items that mapped onto TPOCS-RS subscales were considered for exclusion from subsequent analyses due to poor interrater reliability (ICC < .40): "Explores Past," "Recruits Others." Both items evidenced poor interrater reliability across two of the three conditions, so these items were not included in the TPOCS-RS Psychodynamic and Family subscales for subsequent analyses.

Construct Validity

The next analyses evaluated the construct validity of scores for the TPOCS-RS Anxiety subscale and Consultation Record scales (Consultation Standard, Consultation Modular Anxiety). The construct validity of scores for TPOCS-RS Anxiety subscale was evaluated separately within each condition. Evidence of construct validity was evaluated by examining the magnitude and pattern of correlations among scores on the observer-rated TPOCS-RS subscales (Anxiety, Non-Standard/Non-Modular, Psychodynamic, Family, Client-Centered), consultation records (Consultation Standard, Consultation Modular Anxiety), and the observer-rated alliance measure (TPOCS-A). We hypothesized that within each condition, scores on the condition specific TPOCS-RS subscale (Anxiety) would demonstrate evidence of convergent validity via a strong correlation with the corresponding Consultation Record scale (Standard, Modular Anxiety). We hypothesized that within each condition the TPOCS-RS condition specific subscale would demonstrate evidence of discriminant validity via small to moderate correlations with scores on the remaining TPOCS-RS subscales (Non-Standard/Modular, Psychodynamic, Family, and Client-Centered) and the TPOCS-A (Hogue et al., 2008). The correlations were interpreted following Rosenthal and Rosnow's (1984) guidelines: *r* is a "small" effect when at least .10, "medium" effect when at least .24, and a "large" effect when at least .37.

Standard condition.—As seen in Table 4, there was a moderate correlation between the TPOCS-RS Anxiety subscale and the Consultation Standard scale ($r = .36$). Correlations between scores on the TPOCS-RS Anxiety subscale and the remaining TPOCS-RS subscales (Non-Standard, Psychodynamic, Family, Client-Centered) and the TPOCS-A were all small to moderate and ranged from $r = -.30$ to .20. Similarly, correlations between scores on the Consultation Standard scale and the remaining TPOCS-RS subscales and TPOCS-A were small to moderate in magnitude, ranging from $r = -.31$ to .24. Comparisons indicated that the correlation between the TPOCS-RS Anxiety subscale and the Consultation Standard scale ($r = .36$) was significantly higher than (a) the correlation between the TPOCS-RS Non-Standard subscale and the Consultation Standard scale, $r = .04$, $z = 4.64$, $p < .001$, (b) the mean of the absolute value of the correlations between the TPOCS-RS Anxiety subscale and the remaining subscales ($M\,r = .16$, $SD = .12$; $z = 2.88$, $p = .002$), and (c) the mean of the absolute value of the correlations between the Consultation Standard scale and the remaining subscales ($M\,r = .12$, $SD = .14$; $z = 3.42$, $p < .001$). In all, these findings support the convergent and discriminant validity of the TPOCS-RS Anxiety subscale and the Consultation Standard scale.

Modular condition.—As seen in Table 5, there was a large correlation between the TPOCS-RS Anxiety subscale and the Consultation Modular Anxiety scale ($r = .46$). The correlations between scores on the TPOCS-RS Anxiety subscale and the remaining TPOCS-RS subscales (Non-Modular, Psychodynamic, Client-Centered, Family) and the TPOCS-A were small to medium, ranging from $r = -.25$ to .21. Similarly, correlations between scores on the Consultation Modular Anxiety scale and the remaining TPOCS-RS subscales and the TPOCS-A were small to medium and ranged from $r = -.24$ to .10. The correlation between scores on the TPOCS-RS Anxiety subscale and the Consultation Modular Anxiety scale ($r = .46$) was significantly higher than (a) the correlation between the TPOCS-RS Non-Modular subscale and the Consultation Modular Anxiety scale, $r = -.10$, $z = 4.72$, $p < .001$, (b) the mean of the absolute value of correlations between the TPOCS-RS Anxiety subscale and the remaining subscales ($M\,r = .13$, $SD = .11$; $z = 5.13$, $p < .001$), and (c) the mean of the absolute value of correlations between the Consultation Modular Anxiety scale and remaining subscales ($M\,r = .11$, $SD = .08$; $z = 5.46$, $p < .001$). These findings support the convergent and discriminant validity of the TPOCS-RS Anxiety subscale and Consultation Modular Anxiety scale.

Variance Components Analysis

A series of variance components analyses was conducted to ascertain whether targets of measurement were associated with variation in the scores on the TPOCS-RS subscales within the (a) Standard and Usual Care conditions and (b) Modular and Usual Care conditions. Variance components analysis partitions the total variance among scores into reliable sources of variance (e.g., youth, clinician). Mixed models procedures were employed to account for the nested design in SAS/STAT Software 9.4 (see Barber et al., 2004; McLeod et al., 2015). Restricted maximum likelihood estimation was used to calculate variance components for the following random factors: (a) Condition; (b) Clinician (nested in condition); (c) Youth (nested in condition, clinician); (d) Time (nested in youth, clinician, condition); and (e) Coder. Each factor represents a potential source of variance

in treatment integrity (see e.g., Barber et al., 2004; McLeod et al., 2015). "Condition" refers to the influence of the groups (Standard and Usual Care; Modular and Usual Care) on variation in the TPOCS-RS subscale scores; "clinician" represents differences across clinicians on a TPOCS-RS subscale score; "youth" refers to differences across youth on TPOCS-RS subscale scores; "time in treatment" reflects the influence time on treatment has on each TPOCS-RS subscale score; "coder" reflects differences across coder ratings on TPOCS-RS subscales. The analyses were run separately for Standard (and Usual Care) and Modular (and Usual Care). We hypothesized that condition would account for the highest proportion of the variance on the TPOCS-RS Anxiety subscale as the conditions are expected to differentiate from non-manualized usual care in the techniques contained in the two programs. Each effect was entered as a random effect, and the estimates of variance were transformed into proportions of variance based on estimates of the total variance. A separate analysis for each TPOCS-RS subscale score was run within each condition. We also evaluated the discriminant validity of the TPOCS-RS Anxiety subscale by ascertaining if the subscale scores could identify expected differences between each treatment (Standard, Modular) and Usual Care. To evaluate discriminant validity, we produced adjusted least square means (LSMs) scores for the TPOCS-RS Anxiety subscale score within each condition derived from the mixed-model analysis used for the variance components. This process produces subscale scores that are corrected for the influence of other variables (i.e., condition, clinician, youth, time in treatment, coder). For these analyses, we recalculated the TPOCS-RS Anxiety subscale using the highest scored item. This procedure is intended to provide a more accurate estimate of the dose delivered to youth (Smith et al., 2017). We hypothesized that the Anxiety subscale would have significantly higher scores in each condition (Standard or Modular) than in Usual Care.

Standard condition.—As seen in Table 6, condition accounted for variance in the TPOCS-RS Anxiety subscale (.58 or 58%). A smaller proportion of variance in the remaining TPOCS-RS subscales (Non-Standard, Psychodynamic, Family, Client-Centered) were accounted for by condition (range from .02 to .09). This suggests that scores on the TPOCS-RS Anxiety subscale varied across the Standard and Usual Care conditions, and to a lesser extent varied across the other subscales. Time in treatment accounted for a substantial amount of variation in all TPOCS-RS subscale scores (range from .19 to .68), suggesting that all subscale scores varied over treatment. Clinician accounted for variation in the TPOCS-RS Non-Standard, Psychodynamic, Family, and Client-Centered subscales (range from <.01 to .24), indicating that the use of some techniques likely varied across clinicians. Youth and coder did not account for more than 3% of the total variation in scores on any TPOCS-RS subscales. Since condition accounted for a significant proportion of the variance in the TPOCS-RS Anxiety subscale a follow up contrast was conducted. As hypothesized, the score on the TPOCS-RS Anxiety subscale was significantly higher in Standard (M = 4.33) than in Usual Care (M = 1.55), $t(37) = 30.78$, $p < .001$. Overall, these findings suggest that condition and time account for the most variation in scores on the TPOCS-RS Anxiety subscale.

Modular condition.—As seen in Table 7, a substantial proportion of variance in the TPOCS-RS Modular subscale was accounted for by condition (.37), and a

smaller proportion of variance in the remaining TPOCS-RS subscales (Non-Modular, Psychodynamic, Family, Client-Centered) were accounted for by condition (range from <.01 to .05). This suggests that scores on the TPOCS-RS Anxiety subscale varied across Modular and Usual Care. Across all TPOCS-RS subscales, time in treatment accounted for variation in scores (range from .23 to .46), suggesting that the scores varied over treatment. Clinician accounted for variation in the subscales (range from .06 to .24), with the highest proportion of variation accounted for in the Client-Centered (.24) subscale. Youth did not account for more than 4% of the total variation in subscale scores, and coder did not account for significant variation in any subscale scores. Since condition accounted for a significant proportion of the variance in the TPOCS-RS Anxiety subscale, we conducted follow up contrasts. As hypothesized, LSM indicated that the scores on the TPOCS-RS Anxiety subscale was significantly higher in Modular (M = 3.37) than in Usual Care (M = 1.56), $t(31) = 18.97$, $p < .001$. Overall, condition and time in treatment account for the highest proportion of variation in the TPOCS-RS Anxiety subscale.

Discussion

The main goal of this study was to evaluate the potential of the TPOCS-RS to estimate protocol adherence for two treatment programs evaluated in a randomized effectiveness trial. To achieve this goal, the interrater reliability, construct validity, and discriminant validity of the TPOCS-RS item and subscale scores were evaluated. Findings indicate that the specific TPOCS-RS cognitive and behavioral items that map onto the protocol content of the two treatment programs evidenced adequate interrater reliability within each treatment condition (ICC > .40; Cicchetti, 1994). Scores on the TPOCS-RS Anxiety subscale, designed to map onto the content of both treatments, demonstrated evidence of construct and discriminant validity. Notably, scores on the TPOCS-RS Anxiety subscale converged with consultant-report measures designed to assess protocol adherence for each treatment program. Together, these findings indicate that the TPOCS-RS shows promise for estimating protocol adherence and support the score validity of consultant-reported records used to estimate protocol adherence.

When based on the full sample, the average interrater reliability of the TPOCS-RS items (M ICC = .67) was consistent with mean values reported in previous studies (M ICC range = .61 to .84; e.g., McLeod & Weisz, 2010; McLeod et al., 2015). The mean interrater item-level reliability was consistent within each condition, though usual care had twice the number of items falling in the “poor” interrater reliability range (ICCs < .40; Cicchetti, 1994). The TPOCS-RS cognitive and behavioral items that mapped onto the anxiety-focused content of the two treatment protocols evidenced adequate interrater reliability in the Standard and Modular conditions (all ICCs > .40). Low base rates appear to account for the poor interrater reliability evidenced by a few items, particularly in usual care, a pattern that has been observed in previous studies (e.g., McLeod et al., 2015). In this study, we decided to drop items that demonstrated low interrater reliability (ICC < .40) across two of the three conditions as this pattern indicates the coders did not agree on how to score the item. We did not, however, drop items that evidenced low interrater reliability in a single condition due to low base rates (e.g., low rates of exposure in usual care). In the present study, the pattern of interrater reliability across conditions underscores the importance of calculating interrater

reliability at the planned level of analysis (e.g., single condition, multiple conditions, or total sample). Moreover, our findings indicate that researchers should consider if items will evidence adequate variability in a sample when selecting TPOCS-RS items for use in a study. Overall, our findings indicate that the TPOCS-RS cognitive and behavioral items that mapped onto the anxiety-focused content of the two treatment programs were reliably scored in both conditions.

The magnitude and pattern of correlations within each condition support the construct validity of scores on the TPOCS-RS Anxiety subscale. The TPOCS-RS Anxiety subscale evidenced strong correlations with the consultant-report records and demonstrated smaller correlations with other aspects of treatment delivery and the alliance. The correlations between the observer- and consultant-report measures is consistent with what has been observed for these informants within the literature on adherence measurement (e.g., Dennhag et al., 2012; Ward et al., 2013). Furthermore, the correlations between the TPOCS-RS Anxiety subscale and the other TPOCS-RS subscales and the alliance are consistent with previous research evaluating the association between protocol adherence, treatment delivery, and alliance measures (Hogue et al., 2008; McLeod et al., 2015; Southam-Gerow et al., 2016).

Findings from the variance components analysis further supported the score validity of the TPOCS-RS Anxiety subscale. In both groups, condition accounted for the highest proportion of systematic variation within the TPOCS-RS Anxiety subscale. This is consistent with what would be expected given the training and consultation provided to clinicians in the Standard and Modular conditions (Barber et al., 2004; McLeod et al., 2015). Almost half of the variance in the TPOCS-RS Anxiety subscale in the Standard condition was accounted for by condition, whereas a little over a third of the variance in the Anxiety subscale was accounted for by condition in the Modular condition. This difference may be related to the fact that the Anxiety subscale did not reflect all of the content from the MATCH protocol, since the program contains CBT for anxiety, CBT for depression, and parent training techniques for conduct problems. Follow up analyses indicated that the TPOCS-RS Anxiety subscale scores were significantly higher in both conditions compared to usual care. This indicates that the TPOCS-RS Anxiety subscale can discriminate between conditions and is consistent with the independent adherence ratings conducted in the Child STEPS Multisite Trial (see Weisz et al., 2012). Time in treatment, clinician, and to a lesser extent youth accounted for variation in the scores on the TPOCS-RS Anxiety subscale in both conditions, suggesting that the scores vary over treatment. More broadly, these findings are consistent with previous research evaluating the facets that account for systematic variation in TPOCS-RS subscales (McLeod et al., 2015).

Overall, our findings indicate that the TPOCS-RS items and subscales may be used to estimate protocol adherence. This suggests that it may be possible to combine the TPOCS-RS items to assess protocol adherence for specific treatment programs. To our knowledge, this is the first study to demonstrate that a “generic” measure that is neither problem nor treatment specific can estimate protocol adherence. Considered together with previous support for the TPOCS-RS as a differentiation measure (see McLeod et al., 2015), it may be possible to use the TPOCS-RS to estimate both protocol adherence and differentiation.

Treatment integrity measurement can require significant resources, so the ability to employ one measure to assess adherence and differentiation could be more efficient than using multiple measures.

Our findings also support the score validity of the self-report consultation records, which pushes back against the common viewpoint that observer-report measures should be considered a gold-standard. This narrative is built, in part, on previous findings that indicate observer- and self-report measures evidence poor correspondence (Chapman et al., 2013; Hurlburt et al., 2010). Our findings add to a small group of studies that demonstrate the potential of consultant-report measures to estimate protocol adherence (see Dennhag et al., 2012; Ward et al., 2013). Efforts to improve the accuracy of self-report integrity measures have shown promise (Fallon et al., 2018). Evidence suggests that correspondence may vary depending on the informant (i.e., consultant vs clinician vs client; Chapman et al., 2013) or the type of technique being reported on (e.g., family vs CBT; Hogue et al., 2015). Building on these findings may ultimately provide more cost-effective methods for collecting adherence data.

We acknowledge that the TPOCS-RS does not represent a one size fits all solution to estimating protocol adherence. The TPOCS-RS was not designed to assess techniques found in a particular treatment program (e.g., Coping Cat; Kendall & Hedtke, 2006) or for a specific problem type (e.g., anxiety). The generic focus of the TPOCS-RS means that the items will not precisely fit the techniques defined within a treatment protocol. The generic focus also means that coding does not distinguish between when a clinician addresses cognitive distortions related to depression versus anxiety. This lack of precision means that the TPOCS-RS is not a good fit for certain types of questions that an adherence measure developed for a specific treatment protocol might be able to answer, such as questions addressed in pilot feasibility trials (e.g., were the techniques of a new treatment delivered as designed; Carroll & Nuro, 2002).

Five new TPOCS-RS items were introduced in this study. Four items demonstrated evidence of adequate interrater reliability, with the exception of “Therapy Engagement.” It is important to note that the base rate of “Therapy Engagement” was extremely low, likely contributing to the poor interrater reliability. It is beyond the scope of this study to report on the score validity of these items, which will require evaluation in future research.

A few limitations of the current research bear mentioning. First, though the TPOCS-RS shows promise for assessing protocol adherence and differentiation, it does not assess competence. Competence of delivery is often considered a critical and correlating element of integrity, which would require additional assessment on top of the TPOCS-RS. Second, the TPOCS-RS relies on a rigorous training process that takes about six weeks to complete by coders working 20 hours per week. In addition, the number of items on the TPOCS-RS can be considered a strength in that the TPOCS-RS is broadly applicable; at the same time, this may also be a limitation to its use in that it makes it more difficult to learn. Thus, the TPOCS-RS may be difficult to use outside of funded research projects. Third, our findings only speak to the potential of the TPOCS-RS to estimate protocol adherence, not to provide a complete picture of treatment integrity. Other important ways of estimating integrity exist,

such as practice sequencing (i.e., the ordering of techniques over treatment; Park et al., 2015) and consultant recommendations (i.e., expert recommendations regarding techniques to delivery; Regan et al., 2019). Finally, though averaging scores on a scale or subscale gives an estimate of dose and is a common practice in the field (e.g., Hogue et al., 2008; Southam-Gerow et al., 2016), this approach does not determine how many of the techniques from a treatment protocol were delivered.

While the psychometric properties of the TPOCS-RS demonstrated promise as an approximation of protocol adherence, future research should focus on extending this application across a number of factors, including diverse treatment modalities, problem areas, populations, and settings. These investigations may build confidence that the TPOCS-RS subscales measure therapeutic content similar to measures designed to estimate protocol adherence for a single treatment model. More research may additionally provide a stronger evidence base for estimating the precision of the TPOCS-RS as an approximation of adherence across these factors, and thus an indication of the potentially generalizability of TPOCS-RS applications.

Acknowledgments

Preparation of this article was supported in part by a grant from the National Institute of Mental Health Grant (RO1 MH086529; McLeod & Southam-Gerow). A portion of Dr. Martinez's work on this manuscript was supported by funding from the National Institute of Mental Health (2T32MH073517).

References

Achenbach TM, & Rescorla LA (2000). Manual for the ASEBA preschool forms and profiles (Vol. 30). Burlington, VT: University of Vermont, Research center for children, youth, & families.

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders: DSM-IV-TR. Washington, DC: American Psychiatric Association.

Barber JP, Foltz C, Crits-Christoph P, & Chittams J. (2004). Therapists' adherence and competence and treatment discrimination in the NIDA collaborative cocaine treatment study. Journal of Clinical Psychology, 60(1), 29–41. [PubMed: 14692007]

Barkley RA (2013). Defiant children: A clinician's manual for assessment and parent training. Guilford Press.

Brookman-Frazee LI, Taylor R, & Garland AF (2010). Characterizing community-based mental health services for children with autism spectrum disorders and disruptive behavior problems. Journal of Autism and Developmental Disorders, 40(10), 1188–1201. [PubMed: 20204690]

Carroll KM, & Nuro KF (2002). One size cannot fit all: A stage model for psychotherapy manual development. Clinical Psychology: Science and Practice, 9, 396–406.

Chapman JE, McCart MR, Letourneau EJ, & Sheidow AJ (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(4), 674–680. [PubMed: 23668668]

Chorpita BF, Bernstein A, Daleiden EL, & the Research Network on Youth Mental Health. (2008). Driving with roadmaps and dashboards: Using information resources to structure the decision models in service organizations. Administration and Policy in Mental Health and Mental Health Services Research, 35, 114–123. [PubMed: 17987376]

Chorpita BF, & Weisz JR (2005). Modular Approach to Therapy for Children with Anxiety, Depression, and Conduct Problems (MATCH-ADC). Unpublished Treatment Manual.

Cicchetti D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment, 6, 284–290.

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 & Adolescent Psychiatry, 29(4), 1–15.

Cox JR, Martinez RG, & Southam-Gerow MA (2019). Treatment integrity in psychotherapy research and implications for the delivery of quality mental health services. Journal of Consulting and Clinical Psychology, 87(3), 221–233. [PubMed: 30589351]

Dennhag I, Gibbons MBC, Barber JP, Gallop R, & Crits-Christoph P. (2012). How many treatment sessions and patients are needed to create a stable score of adherence and competence in the treatment of cocaine dependence? Psychotherapy Research, 22(4), 475–488. [PubMed: 22449079]

Fallon LM, Sanetti LMH, Chafouleas SM, Faggella-Luby MN, & Briesch AM (2018). Direct training to increase agreement between teachers' and observers' treatment integrity ratings. Assessment for Effective Interventions, 43(4), 196–211.

Fjermestad KW, McLeod BD, Heiervang ER, Havik OE, Öst LG, & Haugland BSM (2012). Factor structure and validity of the Therapy Process Observational Coding System for Child Psychotherapy-Alliance scale. Journal of Clinical Child and Adolescent Psychology, 41(2), 246–254. [PubMed: 22417197]

Garland AF, Accurso EC, Haine-Schlagel R, Brookman-Frazee L, Roesch S, Zhang JJ (2014). Searching for elements of evidence-based practices in children's usual care and examining their impact. Journal of Clinical Child and Adolescent Psychology, 43(2), 201–215. [PubMed: 24555882]

Garland AF, Brookman-Frazee L, Hurlburt MS, Accurso EC, Zoffness RJ, Haine-Schlagel R, Ganger W. (2010). Mental health care for children with disruptive behavior problems: A view inside therapists' offices. Psychiatric Services, 61(8), 788–795. [PubMed: 20675837]

Haugland BSM, Haaland ÅT, Baste V, Bjaastad JF, Hoffart A, Rapee RM, Raknes S, Himle JA, Husabø E, & Wergeland GJ (2020). Effectiveness of brief and standard school-based cognitive-behavioral interventions for adolescents with anxiety: A randomized noninferiority study. Journal of the American Academy of Child and Adolescent Psychiatry, 59(4), 552–564. [PubMed: 31926224]

Hogue A, Dauber S, Chinchilla P, Fried A, Henderson C, Inclan J, Reiner RH, & Liddle HAS (2008). Assessing fidelity in individual and family therapy for adolescent substance abuse. Journal of Substance Abuse Treatment, 35(2), 137–147. [PubMed: 17997268]

Hogue A, Dauber S, Lichvar E, Bobek M, & Henderson CE (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, 42(2), 229–243. [PubMed: 24711046]

Hogue A, Liddle HA, & Rowe C. (1996). Treatment adherence process research in family therapy: A rationale and some practical guidelines. Psychotherapy: Theory, Research, Practice, Training, 33(2), 332–345.

Hurlburt MS, Garland AF, Nguyen K, & Brookman-Frazee L. (2010). Child and family therapy process: Concordance of therapist and observational perspectives. Administration and Policy in Mental Health, 37(3), 230–244. [PubMed: 19902347]

Kendall PC, & Hedtke KA (2006). Cognitive-behavioral therapy for anxious children: Therapist manual (3rd ed.). Workbook Publishing.

Liber JM, McLeod BD, Van Widenfelt BM, Goedhart AW, van der Leeden AJ, Utens EM, & Treffers PD (2010). Examining the relation between the therapeutic alliance, treatment adherence, and outcome of cognitive behavioral therapy for children with anxiety disorders. Behavior Therapy, 41(2), 172–186. [PubMed: 20412883]

Marques L, Valentine SE, Kaysen D, Mackintosh M-A, …Wiltsey-Stirman S. (2019). Provider fidelity and modifications to cognitive processing therapy in a diverse community health clinic: Associations with clinical change. Journal of Consulting and Clinical Psychology, 87(4), 357–369. [PubMed: 30883163]

McLeod BD, Smith MM, Southam-Gerow MA, Weisz JR, & Kendall PC (2015). Measuring treatment differentiation for implementation research: The Therapy Process Observational Coding System for Child Psychotherapy Revised Strategies scale. Psychological Assessment, 27(1), 314–325. [PubMed: 25346995]

McLeod BD, & Weisz JR (2005). The Therapy Process Observational Coding System-Alliance Scale: Measure characteristics and prediction of outcome in usual clinical practice. Journal of Consulting and Clinical Psychology, 73, 323–333. [PubMed: 15796640]

McLeod BD, & Weisz JR (2010). The therapy process observational coding system for child psychotherapy strategies scale. Journal of Clinical Child and Adolescent Psychology, 39(3), 436–443. [PubMed: 20419583]

Park AL, Chorpita BF, Regan J, Weisz JR, & The Research Network on Youth Mental Health (2015). Integrity of evidence-based practice: Are providers modifying practice content or practice sequencing? Administration and Policy in Mental Health, 42, 186–196. [PubMed: 24841745]

Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, Griffey R, & Hensley M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. Administration and Policy in Mental Health, 38(2), 65–76. [PubMed: 20957426]

Regan J, Park A, & Chorpita B. (2019). Choices in treatment integrity: Considering the protocol and consultant recommendations in child and adolescent therapy. Journal of Clinical Child and Adolescent Psychology, 48(sup1), S79–S89. [PubMed: 27704985]

Rosenthal R, & Rosnow RL (1984). Essentials of behavioral research: Methods and data analysis. McGraw-Hill.

Schoenwald SK, Garland AF, Chapman JE, Frazier SL, Sheidow AJ, Southam-Gerow MA (2011). Toward the effective and efficient measurement of implementation fidelity. Administration and Policy in Mental Health, 38(1), 32–43. [PubMed: 20957425]

Shrout PE, & Fleiss JL (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86(2), 420–428. [PubMed: 18839484]

Smith MM, McLeod BD, Southam-Gerow MA, Jensen-Doss A, Kendall PC, & Weisz JR (2017). Does the delivery of CBT for youth anxiety differ across research and practice settings? Behavior Therapy, 48(4), 501–516. [PubMed: 28577586]

Southam-Gerow MA, McLeod BD, Arnold CC, Rodríguez A, Cox JR, Reise SP, Bonifay WE, Weisz JR, & Kendall PC, (2016). Initial development of a treatment adherence measure for cognitive–behavioral therapy for child anxiety. Psychological Assessment, 28(1), 70–80. [PubMed: 26011477]

Southam-Gerow MA, Weisz JR, Chu BC, McLeod BD, Gordis EB, & Connor-Smith JK (2010). Does cognitive behavioral therapy for youth anxiety outperform usual care in community clinics? An initial effectiveness test. Journal of the American Academy of Child & Adolescent Psychiatry, 49(10), 1043–1052. [PubMed: 20855049]

Ward AM, Regan J, Chorpita BF, Starace N, Rodriguez A, Okamura K, Daleiden EL, Bearman SK, & Weisz JR (2013). Tracking evidence based practice with youth: validity of the MATCH and standard manual consultation records. Journal of Clinical Child & Adolescent Psychology, 42, 44–55. [PubMed: 22809135]

Weisz JR, Chorpita BF, Palinkas LA, Schoenwald SK, Miranda J. …Mayberg S. (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. [PubMed: 22065252]

Weisz JR, Weersing VR, Valeri SM, & McCarty CA (1999). Therapist's Manual PASCET: Primary and secondary control enhancement training program. Los Angeles: University of California.

Weller EB, Weller RA, Rooney MT, & Fristad MA (1999a). CHIPS-Children's Interview for Psychiatric Syndromes. Washington, DC: American Psychiatric Press.

Weller EB, Weller RA, Rooney MT, & Fristad MA (1999b). Children's Interview for Psychiatric Syndromes-Parent Version (P-ChIPS). Washington, DC: American Psychiatric Association.

Highlights

- A single measure can assess both protocol adherence and differentiation
- Observer- and consultant-report adherence measures demonstrated convergence
- Findings offer new ways of assessing treatment integrity in effectiveness research

Table 1

Youth Descriptive Data and Comparisons Across Conditions

Variable	M (*SD*) or %			*F* or X^2	*P*
	Standard (*n* = 22)	Modular (*n* = 16)	Usual Care (*n* = 17)		
Age	9.77 (1.51)	9.94 (1.88)	10.00 (1.87)	0.91	.91
Sex					
Female	50.0%	43.7%	41.2%	0.33	.85
Male	50.0%	56.3%	58.8%		
Race/Ethnicity				10.56	.39
White	72.7%	43.8%	41.2%		
Black	--	12.5%	5.9%		
Asian American	4.5%	--	--		
Latinx	4.5%	--	5.9%		
Multiracial	18.2%	37.5%	41.2%		
Other	--	6.3%	5.9%		
CBCL					
Total	65.27 (7.49)	63.62 (10.39)	66.35 (5.23)	0.50	.61
Internalizing	70.00 (6.72)	69.56 (9.33)	68.82 (5.68)	0.13	.88
Externalizing	59.00 (11.28)	55.06 (11.64)	60.18 (8.76)	1.04	.36
Family Income				3.77	.15
Up to 60k per year	54.5%	31.30	70.6%		
Number of Sessions	21.91 (11.17)	20.69 (6.15)	20.87 (11.95)	0.08	.92
Weeks in Treatment	32.05 (13.50)	30.38 (7.71)	38.06 (30.43)	0.73	.49
Session Length	40.49 (13.53)	43.34 (12.27)	40.65 (10.70)	4.22	.02
Number of Coded Sessions	16.32 (8.71)	15.25 (5.98)	11.35 (7.79)	2.10	.13

Note. Standard = Standard manualized treatment condition; Modular = Modular manualized treatment condition; Usual Care = Usual Care treatment condition; CBCL = Child Behavior Checklist.

Table 2

Clinician Descriptive Data and Comparisons Across Conditions

Variable	M (*SD*) or %			*F* or X^2	*P*
	Standard (*n* = 16)	Modular (*n* = 10)	Usual Care (*n* = 13)		
Age	43.56 (9.96)	35.20 (6.81)	40.00 (9.18)	2.66	.08
Sex					
Female	81.2%	80.0%	76.9%	0.09	.96
Male	18.8%	20.0%	23.1%		
Race/Ethnicity				8.99	.34
White	50.0%	40.0%	61.5%		
Black	18.8%	10.0%	--		
Asian American	12.5%	40.0%	23.1%		
Multiracial	6.3%	--	--		
Other	--	10.0%	--		
Degree Type				4.90	.56
MSW	37.5%	50.0%	38.5%		
MA Psych	31.3%	20.0%	23.1%		
PsyD/PhD	6.3%	10.0%	30.8%		
Other	25.0%	20.0%	7.7%		
License	62.5%	60.0%	23.1%	5.70	.22
Years of Experience	7.17 (7.75)	5.25 (4.83)	4.69 (5.34)	0.57	.57
Theoretical Orientation				3.22	.92
CBT	31.3%	40.0%	30.8%		
Psychodynamic	25.0%	20.0%	15.4%		
Family Systems	6.3%	--	7.7%		
Eclectic	31.3%	40.0%	30.8%		
Other	6.3%		15.4%		

Note. Standard = Standard manualized treatment condition; Modular = Modular manualized treatment condition; Usual Care = Usual Care treatment condition; MSW = Master of Social Work; PsyD = Doctor of Psychology; PhD = Doctor of Philosophy; CBT = cognitive behavioral therapy.

Table 3

TPOCS-RS Item Descriptions and Interrater Reliability

	Standard		Modular		Usual Care		Total
TPOCS-RS Item and Definition	**M(SD)**	**ICC**	**M(SD)**	**ICC**	**M(SD)**	**ICC**	**ICC**
Cognitive							
*__Cognitive Education:__ Teaches cognitive model/Identifies how the cognitive model applies to an aspect of the client's life.	2.51(1.53)	.84	1.53(.96)	.74	1.15(.43)	.78	.85
*__Cognitive Distortion:__ Teaches/encourages client to identify/restructure cognitive distortions.	2.23(1.44)	.82	1.48 (.84)	.75	1.22(.60)	.86	.83
*__Coping Skills:__ Teaches/encourages client to use coping skills (e.g., problem-solving skills).	2.11(1.42)	.82	1.25(.50)	.53	1.18(.44)	.33	.82
Behavioral							
__Functional Analysis of Behavior:__ Performs functional analysis/teaches A-B-C model.	1.10(.32)	.49	1.16(.44)	.61	1.09(.22)	.15	.51
*__Relaxation Interventions:__ Teaches/encourages client to use relaxation.	1.60(1.20)	.93	1.17(.66)	.88	1.12(.46)	.84	.92
*__Respondent Interventions:__ Develops a fear hierarchy and/or performs an exposure.	2.27(1.61)	.87	2.68(1.71)	.86	1.05(.19)	.24	.88
*__Operant Strategies-Child:__ Teaches principles of operant interventions, sets up operant system, or employs operant intervention.	1.64(.68)	.54	1.43(.71)	.66	1.09(.22)	.57	.63
__Skill Building:__ Interventions that focus on building behavioral skills (e.g., social skills).	1.10(.32)	.49	1.16(.44)	.61	1.09(.22)	.15	.51
__Behavioral Activation:__ Teaches/demonstrates relation between pleasant activities and mood improvement or assigns participation in a pleasant event to improve mood.	1.60(1.20)	.93	1.17(.66)	.88	1.12(.46)	.84	.92
__Monitoring:__ The therapist elicits measurements of a client's mood, functioning, or experience, or encourages a client to monitor his/her mood.	2.27(1.61)	.87	2.68(1.71)	.86	1.05(.19)	.24	.88
__Modeling:__ Teaches specific skills using observational learning methods.	1.64(.68)	.54	1.43(.71)	.66	1.09(.22)	.57	.63
Psychodynamic							
__Addresses Transference:__ Discusses or interprets the client's interaction with the therapist.	1.00(.06)	.80	1.01(.10)	.00	1.03(.19)	.68	.55
__Explores Past:__ Discusses client's past experiences.	1.01(.08)	00	1.00(.03)	.00	1.03(.18)	.49	.35
__Addressing Client Resistance:__ Identifies/processes client's resistance to therapy and/or resistance to change.	1.01(.08)	−.01	1.04(.25)	.81	1.05(.25)	.31	.54
__Interpretation:__ Comments on client behavior and/or relates that behavior to an aspect of the client's characteristics, general functioning, and/or past experiences.	1.05(.19)	.52	1.07(.23)	.48	1.22(.44)	.37	.44
Family							
__Targets Other Participants:__ Participants other than the target child are asked to modify their affect, behavior, cognitions.	1.50(.19)	.93	2.09(1.23)	.81	1.90(1.22)	.81	.87
__Recruits Others:__ Tries to recruit/retain parents and other family members for future sessions.	1.05(.23)	.54	1.07(.23)	.24	1.05(.16)	.15	.38
__Parenting Style:__ Helps and/or encourages parents to modify their parenting practices.	1.05(.36)	84	1.13(.63)	.92	1.05(.24)	.51	.86
^__Operant Strategies-Parent:__ Teaches principles of operant interventions, sets up operant system, or employs operant intervention with parents.	1.07(.42)	.80	1.51(1.13)	.88	1.07(.33)	.79	.88

	Standard		Modular		Usual Care		Total
TPOCS-RS Item and Definition	**M(SD)**	**ICC**	**M(SD)**	**ICC**	**M(SD)**	**ICC**	**ICC**
Parenting Skills: Parent-focused interventions designed to improve ability to set limits, ensure adequate supervision, and monitor child behavior (appropriate or inappropriate), activities, and involvements.	1.08(.37)	.74	1.38(.75)	.48	1.14(.45)	.54	.54
Multiparticipant Interactions: Establishes/teaches/discusses in-session interactions.	1.17(.56)	.71	1.45(.80)	.66	1.14(.45)	.79	.74
Family Members' Roles: Therapist teaches or emphasizes how problems may be caused/maintained by family dynamics.	1.15(.54)	.78	1.08(.32)	.18	1.57(.97)	.58	.64
Client-Centered							
Validates Client: Validates client's feelings and/or treatment goals.	1.35(.64)	.67	1.32(.51)	.51	1.64(.79)	.68	.66
Positive Regard: Responds to client in warm and compassionate manner.	2.22(.95)	.54	2.06(.91)	.60	1.89(.89)	.56	.57
Client Perspective: Attempts to understand client's point of view/Probes for client's unique perspective.	2.63(.98)	.46	2.84(.97)	.52	3.92(1.25)	.66	.63
General Items							
Rehearsal: Encourages client to participate in hypothetical enactments.	2.44(1.25)	.68	1.73(1.00)	.67	1.19(.50)	.60	.74
Homework: Assigns and/or reviews homework assignments.	2.55(1.15)	.74	2.07(.98)	.58	1.25(.61)	.75	.76
Play/Art Therapy: Therapist utilizes play or art as a form of therapy.	1.68(1.04)	.78	1.88(1.24)	.87	3.25(1.87)	.87	.87
Encourages Affect: Discusses affect and/or encourages client to express affect.	1.32(.54)	.43	1.17(.32)	.16	1.66(.94)	.69	.58
Session Goals: Establishes/reviews session goals.	2.07(.86)	.45	2.08(.88)	.44	1.36(.54)	.42	.51
Treatment Goals: Establish treatment goals/Encourage client to discuss treatment goals.	1.98(1.18)	.69	2.36(1.19)	.59	1.67(.96)	.71	.67
Previous Themes: Comments on themes from previous sessions/builds on past successes.	1.88(.88)	.45	1.94(1.11)	.69	1.41(.67)	.53	.59
Psychoeducation: Teaches client about general psychological principles (e.g., anxiety).	1.59(.91)	.67	1.92(1.14)	.65	1.44(.71)	.66	.67
Questioning: Therapist asks about general topics.	1.60(.78)	.70	1.76(.75)	.56	2.43(.92)	.38	.59
Self-Disclosure: Therapist provides information about his/her personal life, feelings, experiences.	1.54(.76)	.65	1.26(.57)	.69	1.36(.63)	.66	.67
Advice: Therapist provides the client with precise instructions on how to address a specific issue.	1.08(.28)	.46	1.07(.29)	.43	1.72(.38)	.23	.38
Coaching: Therapist makes efforts to actively direct the client as they practice a specific skill.	1.45(.75)	.54	1.20(.48)	.52	1.05(.25)	.55	.57
Assessment: Therapist gathers data about psychiatric, social, academic, or medical problems.	1.64(.91)	.65	2.08(1.14)	.66	1.92(1.16)	.71	.68
^**Crisis Management:** Therapist helps the client address an urgent or dangerous event.	1.00(.04)	00	1.01(.17)	.89	1.01(.22)	1	.92
^**Case Management:** Therapist performs activities designed to coordinate services designed to supplement treatment.	1.05(.19)	.34	1.11(.35)	.52	1.21(.50)	.50	.50
^**Therapy Engagement:** Addresses poor treatment participation by discussing barriers to attendance or motivational issues.	1.01(.08)	.40	1.01(.08)	NV	1.01(.09)	.02	.17
^**Educational Support:** Exercises designed to assist the youth with specific academic problems such as homework, study skills, or tutoring.	1.03(.20)	.55	1.00(.06)	NV	1.04(.36)	.93	.77

Note. Standard = Standard manualized treatment condition; Modular = Modular manualized treatment condition; Usual Care = Usual Care treatment condition; TPOCS-RS = The Therapy Process Observational Coding System for Child Psychotherapy – Revised Strategies Scale; ICC = intraclass correlation coefficient.

^Denotes a new TPOCS-RS items

Author Manuscript

Author Manuscript

Author Manuscript

* Denotes items are on the TPOCS-RS Anxiety subscale

Table 4

Correlations among the TPOCS-RS Subscales, Consultation Record, and TPOCS-A in the Standard Condition

	2.	3.	4.	5.	6.	7.
1. TPOCS-RS Anxiety	.36**	−.07	.12*	−.30**	−.02	.20**
2. Consultation Standard		.04	−.04	−.31**	−.11	.01
3. TPOCS-RS Non-Standard			.12	.04	.13*	−.01
4. TPOCS-RS Psychodynamic				.09	.21**	.08
5. TPOCS-RS Family					.36**	.02
6. TPOCS-RS Client-centered						.24**
7. TPOCS-A						

Note: TPOCS-RS = The Therapy Process Observational Coding System for Child Psychotherapy – Revised Strategies scale; TPOCS-A = The Therapy Process Observational Coding System for Child Psychotherapy – Alliance scale.

* $p < .05$,

** $p < .01$.

N = 359

Table 5

Correlations among the TPOCS-RS subscales, Consultation Record, and TPOCS-A in the Modular Condition

	2.	3.	4.	5.	6.	7.
1. TPOCS-RS Anxiety	.46**	−.17**	.01	−.25**	.01	.21**
2. Consultation Modular Anxiety		−.10	.03	−.24**	−.06	.10
3. TPOCS-RS Non-Modular			.06	.12	.06	−.11
4. TPOCS-RS Psychodynamic				−.01	.17**	−.42**
5. TPOCS-RS Family					.08	−.25**
6. TPOCS-RS Client-centered						.16*
7. TPOCS-A						

Note: TPOCS-RS = The Therapy Process Observational Coding System for Child Psychotherapy – Revised Strategies scale; TPOCS-A = The Therapy Process Observational Coding System for Child Psychotherapy – Alliance scale.

* $p < .05$,

** $p < .01$.

N = 244

Table 6

Variance Components for TPOCS-RS Subscales in the Standard and Usual Care Conditions

	Variance Components					
TPOCS-RS Subscale	**Condition**	**Clinician**	**Youth**	**Time**	**Coder**	**Residual**
Anxiety	.58	<.01	.03	.26	<.01	.12
Non-Standard	.02	.08	<.01	.32	<.01	.57
Psychodynamic	.09	.09	<.01	.20	<.01	.63
Family	.04	.12	.01	.68	<.01	.14
Client-centered	.06	.24	<.01	.19	.02	.48

Note. TPOCS-RS = The Therapy Process Observational Coding System for Child Psychotherapy – Revised Strategies scale. Variance components estimates represent the portion of variance that is attributed to each source of variance.

Table 7

Variance Components for TPOCS-RS Subscales in the Modular and Usual Care Conditions

	Variance Components					
TPOCS-RS Subscale	**Condition**	**Clinician**	**Youth**	**Time**	**Coder**	**Residual**
Anxiety	.37	.06	<.01	.33	<.01	.23
Non-Modular	<.01	.09	<.01	.32	<.01	.58
Psychodynamic	.04	.06	.02	.28	<.01	.60
Family	<.01	.09	.O6	.46	<.01	.39
Client-centered	.05	.24	.01	.23	.01	.46

Note. TPOCS-RS = The Therapy Process Observational Coding System for Child Psychotherapy – Revised Strategies scale. Variance components estimates represent the portion of variance that is attributed to each source of variance.