Measuring Treatment Differentiation for Implementation Research: The Therapy Process Observational Coding System for Child Psychotherapy Revised Strategies Scale

Bryce D. McLeod,
Virginia Commonwealth University

Meghan M. Smith,
Virginia Commonwealth University

Michael A. Southam-Gerow,
Virginia Commonwealth University

John R. Weisz, and
Harvard University

Philip C. Kendall
Temple University

Abstract

Observational measures to assess implementation integrity (the extent to which components of an evidence-based treatment are delivered as intended) are needed. We evaluated the reliability of the scores and the validity of the score interpretations for the Therapy Process Observational Coding System for Child Psychotherapy – Revised Strategies scale (TPOCS-RS; McLeod, 2010) and assessed the potential of the TPOCS-RS to assess treatment differentiation, a component of implementation integrity. The TPOCS-RS includes five theory-based subscales (Cognitive, Behavioral, Psychodynamic, Client-Centered, Family). Using the TPOCS-RS, coders independently rated 954 sessions conducted with 89 children (*M* age = 10.56, *SD* = 2.00; aged 7–15 years; 65.20% Caucasian) diagnosed with a primary anxiety disorder who received different treatments (manual-based vs. non-manualized) across settings (research vs. practice). Coders produced reliable ratings at the item level (*M* ICC = .76, *SD* = .18). Analyses support the construct validity of the Cognitive and Behavioral subscale scores and, to a lesser extent, the Psychodynamic, Family, and Client-Centered subscale scores. Correlations among the TPOCS-RS subscale scores and between the TPOCS-RS subscale scores and observational ratings of the alliance and client involvement were moderate suggesting independence of the subscale scores. Moreover, the TPOCS-RS showed promise for assessing implementation integrity as the TPOCS-RS subscale scores, as hypothesized, discriminated between manual-guided treatment delivered across research and practice settings and non-manualized usual care. The findings support the potential of the TPOCS-RS Cognitive and Behavioral subscales to assess treatment differentiation

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.
Bryce D. McLeod, Meghan M. Smith, Michael A. Southam-Gerow, Department of Psychology, Virginia Commonwealth University; John R. Weisz, Department of Psychology, Harvard University; Philip C. Kendall, Department of Psychology, Temple University.

in implementation research. Results for the remaining subscales are promising, although further research is needed.

Keywords

Treatment differentiation; implementation; CBT; usual care; child mental health care

Over the past decade, researchers and service agencies have built active collaborations to implement evidence-based treatments (EBTs) in practice settings (Center for the Study and Prevention of Violence, 2010). Implementation science has gained momentum (Aarons, Hurlburt, & Horwitz, 2010; Weisz, Ng, & Bearman, 2014), spurred by data suggesting that it takes approximately 17 years for EBTs to be disseminated into practice settings (Institute of Medicine, 2001). With regard to youth mental health care, implementation research strives to increase the speed and quality of information transmission between the science and practice of EBTs.

Implementation science addresses the factors that influence the translation of science to practice (Aarons et al., 2010), with a central goal being to ascertain how EBTs can produce beneficial, sustainable effects in practice settings. Central to this goal is determining how EBTs can be delivered with integrity across practice settings (Allen, Linnan, & Emmons, 2012; McLeod, Southam-Gerow, Tully, Rodriguez, & Smith, 2013; Schoenwald et al., 2011). Defined as the extent to which the elements of an EBT are delivered according to the treatment model (Schoenwald et al., 2011), implementation integrity is a critical measurement domain.

Few measures exist for characterizing implementation integrity (Schoenwald et al., 2011). It has been proposed that treatment integrity methods (also called treatment fidelity; e.g., Allen et al., 2012; Bellg et al., 2004) developed for treatment and evaluation research can address this measurement need (McLeod et al., 2013; Schoenwald et al., 2011). Researchers have yet to agree on a single definition of treatment integrity (e.g., Bellg et al., 2004; Hagermoser Sanetti, & Kratochwill, 2009; Perepletchikova & Kazdin, 2005; Waltz, Addis, Koerner, & Jacobson, 1993), but efforts to define implementation integrity have contained the following three treatment integrity components (see Allen et al., 2012; McLeod et al., 2013; Schoenwald et al., 2011): (a) *Treatment adherence* – the extent to which interventions considered integral to the treatment model(s) are delivered, (b) *Treatment differentiation* – the extent to which interventions not part of the intended model are delivered, and (c) *Competence* - the quality and responsiveness of treatment delivery. In addition, some have proposed that a fourth treatment integrity component (e.g., Lichstein, Riedel, & Grieve, 1994; called *patient receipt* or *relational elements*), is relevant when assessing implementation integrity, as perfect adherence to a treatment manual may not produce optimal outcomes unless there is a strong alliance and client involvement (Allen et al., 2012; McLeod et al., 2013). Together, it has been proposed that these components can be used to characterize implementation integrity (McLeod et al., 2013).

The lion's share of research on treatment integrity has focused on treatment adherence (Perepletchikova, Treat, & Kazdin, 2007). Though assessing adherence to the treatment

model is important, it is not fully sufficient for implementation research (McLeod et al., 2013). When EBTs are transported to practice settings, it is possible that interventions not found in the EBTs may be delivered, which may influence outcomes (Southam-Gerow et al., 2010; Weisz et al., 2009). Moreover, usual clinical care (UC), a common comparison group in implementation research, encompasses a wide range of therapeutic interventions (Garland et al., 2010; Hurlburt, Garland, Nguyen, & Brookman-Frazee, 2010; McLeod & Weisz, 2010), and may even include some EBTs. For these reasons, it is critical to perform treatment differentiation checks that assess for an array of interventions (McLeod et al., 2013).

At present there are three observational measures that can be used to assess treatment differentiation in implementation research with children and adolescents. The first, the Child Therapy Process Rating System (CTPRS; Hurlburt et al., 2010), characterizes mental health care for youth with externalizing problems. Hurlburt et al. (2010) reported that the interrater reliability (ICC) of the CTPRS items considered together was .69. However, when interrater reliability was calculated at the item level only 76% of the items achieved an ICC above .50. The other measures represent variants of the Therapy Process Observational Coding System for Child Psychotherapy – Strategies scale (TPOCS-S; McLeod, 2001). Originally used to characterize UC (McLeod & Weisz, 2010), the TPOCS-S, along with a modification (Garland et al., 2010), has been used to gauge the contents of UC for youth with internalizing (McLeod & Weisz, 2010) and externalizing (Garland et al., 2010) problems. Thus, most of the research conducted to date with the TPOCS-S has focused on characterizing UC for clinically-referred youths.

As the TPOCS-S encompasses therapeutic interventions from a range of theoretical orientations it may be well-suited to assessment of treatment differentiation in implementation research (McLeod et al., 2013). Preliminary evidence supports this application of the TPOCS-S. The TPOCS-S was used to characterize UC and assess treatment differentiation in two effectiveness trials: (a) Southam-Gerow et al. (2010) used the TPOCS-S to observe 77 sessions from a sample of clinically-referred youths diagnosed with anxiety disorders randomly assigned to cognitive behavioral therapy (CBT) or UC[1], and (b) Weisz et al. (2009) used the TPOCS-S to observe 94 sessions from a sample of clinically-referred youths diagnosed with depressive disorders randomly assigned to CBT or UC. In both studies, the CBT groups scored significantly higher than UC on CBT interventions and significantly lower than UC on psychodynamic interventions. These studies indicate that the TPOCS-S may have applications beyond being a tool for simply characterizing UC by demonstrating potential for using the TPOCS-S to assess treatment differentiation in implementation research.

Although the TPOCS-S has shown initial promise as a treatment differentiation measure, only one study has reported on the reliability and validity of the TPOCS-S scores. McLeod and Weisz (2010) evaluated the preliminary reliability and validity of the TPOCS-S scores

---

[1]Child participants drawn from the Youth Anxiety Study (Southam-Gerow et al., 2010) for this study comprised a portion of the sample from McLeod and Weisz (2010). Though the TPOCS-S was used to code therapy sessions for these studies, this study does not use any of the TPOCS-S data previously reported in these studies.

in a sample of 43 clinically referred youths with internalizing problems observed in 166 UC sessions. The TPOCS-S demonstrated good to excellent interrater reliability (item ICCs ranged from .72– .94) and analyses supported the validity of TPOCS-S subscale scores. However, the study was designed to assess the potential of the TPOCS-S to characterize UC so the psychometric analyses were limited to intercorrelations among the TPOCS-S subscale scores. To examine the potential of the TPOCS-S for assessing treatment differentiation in implementation research, it will be important to (a) assess a broader array of interventions found in EBTs and UC for youth emotional and behavioral problems, and (b) conduct further assessment of psychometric properties in research and practice settings (the contexts most relevant to assessing implementation integrity). The present study addressed these goals.

The TPOCS-S was originally developed to assess a manageable subset of therapeutic interventions, grouped within categories prominent in the youth therapy literature— cognitive, behavioral, psychodynamic, client-centered, and family. Subscale items were specific therapeutic interventions documented in the literature for each of the categories (McLeod & Weisz, 2010). However, subsequent research has suggested that some interventions commonly found in EBTs and/or UC were not included (Bearman, Weisz, & McLeod, 2010; Garland et al., 2010; Hurlburt et al., 2010). For example, adding an item focused on parenting skills (e.g., giving effective commands) to the Family subscale would increase relevance of the system to interventions commonly used with youth externalizing problems (Garland et al., 2010). So, the present study tested an enriched version of the TPOCS-S, called the TPOCS-Revised Strategies scale (TPOCS-RS; McLeod, 2010), which includes 11 additional interventions.

We examined how the TPOCS-RS performs across domains relevant to implementation research. For the TPOCS-RS to be useful for implementation research, coders must be capable of achieving item level reliability across (a) manual-based and non-manualized treatments, and (b) an EBT delivered in different settings. Additionally, as client and therapist characteristics may influence implementation integrity (Barber, Foltz, Crits-Christoph, & Chittams, 2004; Boswell et al., 2013), we evaluated whether targets of measurement that might influence implementation integrity (child, therapist, study group, coder, time in treatment) accounted for systematic variation in scores on the TPOCS-RS subscales. Finally, we tested whether scores on the TPOCS-RS subscales could detect expected differences between an EBT and UC.

We describe the revision of the TPOCS-S and examine the psychometric properties of the TPOCS-RS scores. To achieve study goals, the TPOCS-RS was used to code sessions for youth diagnosed with anxiety disorders from a manual-based CBT delivered in research and practice settings and non-manualized UC. Although previous research has examined the psychometric properties of the TPOCS-S scores (e.g., McLeod & Weisz, 2010), this is the first study to investigate the psychometric properties of the TPOCS-S scores in the context of implementation research with a large, diverse sample.

Method

Participants and Study Sites

Therapy data were collected on 89 child participants from two randomized controlled trials. One (Kendall, Hudson, Gosch, Flannery-Schroeder, & Suveg, 2008) compared the efficacy of individual-CBT (ICBT), family-CBT, and an active control group. Only the ICBT group was used in this study. The second study, called the Youth Anxiety Study (YAS; Southam-Gerow et al., 2010) compared the effectiveness of ICBT (YAS-ICBT) to UC (YAS-UC); both groups were used in this study. The archived data were collected via recorded sessions conducted by separate research groups. Therapy ratings were based on three measures (see below) and the same coders produced scores across all recordings for each measure. Each child participant met the following criteria: (a) a minimum of two audible sessions, and (b) received treatment from a single therapist (see Kendall et al., 2008 and Southam-Gerow et al., 2010 for more details on the participants and procedures).

The 89 child participants aged 7–15 years (*M* age = 10.56, *SD* = 2.00; 65.20% Caucasian; 52.80% male) met diagnostic criteria for a primary anxiety disorder (see Table 1). There were 51 child participants from the Kendall et al. study (2008) and 38 child participants from YAS (see Southam-Gerow et al., 2010). At post-treatment in Kendall et al., 64% of the youth who received ICBT no longer met diagnostic criteria for their primary anxiety disorder. In YAS, at post-treatment 66.70% and 73.70% of youths no longer met diagnostic criteria treatment in the YAS-ICBT and YAS-UC groups, respectively.

There were 45 therapists (55.60% Caucasian; 13.30% male). Therapy in the Kendall et al. study was delivered by clinical psychology doctoral trainees and licensed clinical psychologists (*N* = 16; 12.50% male). Therapists were 81.30% Caucasian, 6.30% Latino, and 6.30% Asian/Pacific Islander (6.30% did not report). In YAS, therapists were clinic employees (*N* = 29) who volunteered to participate and were randomly assigned to groups. Therapists assigned to YAS-ICBT (*N* = 13; 15.40% male) were 53.80% Caucasian, 15.40% Latino, 15.40% Asian/Pacific Islander, and 15.40% mixed/other. Professionally, 30.80% were social workers, 23.10% were Masters level psychologists, 15.30% doctoral level psychologists, and 30.80% reported "other" degree. Therapists assigned to YAS-UC (*N* = 16; 12.50% male) were 43.80% Caucasian, 37.50% Latino, and 12.50% mixed/other (6.30% did not report). Professionally, 25.00% were social workers, 31.20% were Masters level psychologists, 6.30% doctoral level psychologists, and 31.30% reported that they were "other" (6.30% did not report).

Treatments

CBT—Therapists in ICBT and YAS-ICBT delivered *Coping Cat*, a CBT program designed for youth diagnosed with anxiety disorders (Kendall & Hedtke, 2006a, 2006b). *Coping Cat* includes 16 sessions (14 conducted individually with the child; 2 with parents). The program emphasizes anxiety management skills training (e.g., cognitive change, problem-solving) and exposure tasks. Homework is regularly assigned to the youth. In both studies, the therapists were trained using the gold standard in training for randomized controlled trials (i.e., training workshop, treatment manual, and supervision; Sholomskas et al., 2005).

Therapists in both studies reviewed the *Coping Cat* therapist manual, attended a training workshop lead by experts in the Coping Cat program, and participated in regular group supervision with an expert in the Coping Cat program. Adherence to the *Coping Cat*, measured with the Coping Cat Brief Adherence Scale (see Kendall, 1994; Kendall et al., 1997), was found to be high in ICBT and YAS-ICBT. A treatment differentiation check determined that the YAS-ICBT and YAS-UC conditions did not overlap significantly in CBT interventions unique to *Coping Cat*. (See Kendall et al. (2008) and Southam-Gerow et al. (2010) for more details about treatment integrity).

Usual care (UC)—Therapists who provided UC agreed to use the therapeutic interventions they regularly delivered and believed to be effective in their routine practice.

Therapy Process Observational Coding System for Child Psychotherapy– Revised Strategies Scale (TPOCS-RS; McLeod, 2010)—The TPOCS-S (31 items; McLeod, 2001) consists of five theory-based subscales (Cognitive, Behavioral, Psychodynamic, Family, Client-Centered) along with eight items (e.g., *Homework*) that represent interventions that are considered to play a meaningful role in therapy but are not associated with a specific theory-based subscale (Orlinsky, Rønnestad, & Willutzki, 2004; Weersing, Weisz, & Donenberg, 2002). The observer-based methodology used for the TPOCS-S has been widely used to evaluate treatment integrity (e.g., Carroll et al., 2000; Hogue et al., 2008). The scoring strategy involves extensiveness ratings to measure the degree to which therapists use each intervention during a session. In making extensiveness ratings, coders are asked to estimate the extent to which a therapist engages in each intervention during the entire session using a 7-point Likert-type scale with the following anchors: 1 = *not at all*, 4 = *considerably*, and 7 = *extensively*. Extensiveness ratings are comprised of two components: thoroughness and frequency. Thoroughness refers to the depth, complexity, or persistence with which the therapist engages in a given intervention whereas frequency refers to how often a therapist uses an intervention during a session (see Hogue, Liddle, & Rowe, 1996). Both thoroughness and frequency are considered in making a rating; therefore, extensiveness ratings provide quantity, or dosage, information about each intervention. Following Cicchetti (1994), ICCs 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. Previous studies have demonstrated "good" interrater reliability (TPOCS-S subscale ICCs > .60; Cicchetti, 1994) as well as provided preliminary support for the construct validity of the subscale scores (McLeod & Weisz, 2010; Southam-Gerow et al., 2010; Weisz et al., 2009; Wood, Piacentini, Southam-Gerow, Chu, & Sigman, 2006).

Eleven items were added to the TPOCS-S, to form the TPOCS-RS, by drawing from three observational measures: The 27-item PRAC Study TPOCS-S scale (PRAC TPOCS-S; Garland, Brookman-Frazee, & McLeod, 2008), the 38-item TPOCS for Child Group Psychotherapy Scale (TPOCS-G; Bearman et al., 2010), and the 39-item CTPRS (Hurlburt et al., 2010). Each measure underwent extensive development and was designed to characterize therapy for youth with internalizing (Bearman et al., 2010) or externalizing (Garland et al., 2010; Hurlburt et al., 2010) problems. We selected items from these

measures that were not found in the TPOCS-S. Items were added to the Behavioral (*Behavioral Activation*, *Monitoring*, *Skill Building*), Psychodynamic (*Addresses Resistance*), and Family (*Family Members' Roles*, *Parenting Skills*) subscales. Additionally, five general items were added (*Advice, Coaching, Questioning, Self-Disclosure, Assessment*). Adding the general items expanded the number of delivery methods measured by the TPOCS-RS, thus allowing a more thorough assessment of how therapeutic interventions are delivered.

After identifying items to add to the TPOCS-S, we grouped like items with similar content together to form a single TPOCS-RS item. For example, the PRAC TPOCS-S and TPOCS-G both had several items that addressed behavioral parenting skills (e.g., *Monitoring Child Behavior*, *Effective Commands*, *Responding Effectively to Negative Behavior*) that were combined into a single item called *Parenting Skills*. In most cases, only minor modifications were made to the items (e.g., TPOCS-G items were altered for individual- and/or family-focused therapy). The TPOCS-RS (42 items; McLeod, 2010) consisted of five subscales: Cognitive (4 items), Behavioral (9 items), Psychodynamic (5 items), Family (7 items), and Client-Centered (4 items). In addition, there were 13 general items (e.g., *Homework*, *Play Therapy*). See Table 2 for a list and brief description of the TPOCS-RS items.

Measures for Validity Analyses

Therapy Process Observational Coding System for Child Psychotherapy-Alliance scale (TPOCS-A; McLeod & Weisz, 2005)—The TPOCS-A consists of six items that assess affective elements of the client–therapist relationship, and three items that assess client participation in therapeutic activities. Coders observe entire sessions and rate each item on a six-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), internal consistency ranging from .91 to .95 (*M* α = .92), and convergent validity with self-report alliance measures ranging from .48 to .53 (Fjermestad et al., 2012; Liber et al., 2010). Interrater reliability, ICC(2,2), for the TPOCS-A scale in the present sample was .82; internal consistency was .81.

Child Involvement Rating Scale (CIRS; Chu & Kendall, 2004)—The CIRS is a six item measure that assesses aspects of positive and negative involvement. Coders view entire sessions and then rate items on a 6-point scale ranging from 0 (*not at all*) to 5 (*a great deal*). Previously, the CIRS has demonstrated interrater reliability from .76 to .90 and an internal consistency of .73 (Chu & Kendall, 2004, 2009). The interrater reliability, ICC(2,2), for the CIRS scale based upon the full sample was .79; internal consistency of the CIRS was .85.

Coding and Session Sampling Procedures

Coders—Two female doctoral students (*M age* = 25.18, *SD* = 1.77) in clinical psychology (one Asian American and one Caucasian) comprised the team that coded the TPOCS-RS and TPOCS-A. Two doctoral students (*M age* = 27.82, *SD* = 4.27) in clinical psychology (one male; one Caucasian and one Latina) comprised the CIRS coding team.

Coder training—Coder training involved three steps. First, coders received didactic instruction and discussion of the scoring manuals, reviewed sessions with the trainers, and

engaged in exercises designed to expand understanding of each item. Second, coders engaged in coding and results were discussed in weekly meetings. Lastly, each coder independently coded 32 recordings and reliability was assessed against master codes produced by the study principal investigators (initials BDM, MSG). To be certified for independent coding, each coder had to demonstrate "good" reliability on each item (ICC(2, 2) > .59, Cicchetti, 1994).

Assignment and coding of sessions—All sessions for each case were coded except the first and last session as these sessions may contain intake or termination content. Sessions were not rated if (a) shorter than 15 minutes, (b) less than 15 minutes was audible, (c) less than 75% of the dialogue was in English, or (d) missing or damaged. Coding order was determined by random assignment. Each session was double-coded. Coders were naïve to study hypotheses and differences between data sources. Of the 1428 sessions held, 954 (67%) were rated (66% ICBT, 75% YAS-ICBT, 67% YAS-UC). There were no significant differences across groups in terms of the percent of sessions coded.

Data Analyses

Interrater reliability—We investigate the interrater reliability of the TPOCS-RS item and subscale scores. We calculated interrater reliability using ICC (Shrout & Fleiss, 1979). The reliability coefficients represent the model ICC(2, 2) based on a two-way random effects model, which provides an estimate of the ratio of the true score variance to total variance. Thus, these correlations provide a reliability estimate of the mean scores of all coders considered as a whole, and allow for generalizability of the results to other samples.

Discriminant validity—We assessed the discriminant validity of the TPOCS-RS subscale scores. First, we examined the magnitude of the correlations among the five theory-based TPOCS-RS subscales. Because these subscales were designed to measure distinct sets of interventions, we hypothesized that the correlations among the subscale scores would be small to medium in strength (Cohen, 1992; cf. Carroll et al., 2000), except for scores on the Cognitive and Behavioral subscales. We hypothesized that scores on these two subscales would evidence a large correlation based on previous research, though would not be redundant (i.e., too high: < 0.85; cf. McLeod & Weisz, 2010; Weersing et al., 2002). The correlations were interpreted following Cohen's (1992) guidelines: r is a "small" effect if 0.10–0.23, "medium" if 0.24–0.36, and "large" if > 0.36. Subscale scores were produced by calculating the mean score on each of the TPOCS-RS items and then averaging the item scores on each subscale.

We assessed the magnitude of the correlations between the TPOCS-RS subscale scores and scores on observational measures of the alliance (TPOCS-A) and client involvement (CIRS). Given that these measures are designed to assess separate but related therapy processes, we hypothesized that the correlations would be small to medium using Cohen's (1992) standards (cf. Carroll et al., 2000; Hogue et al., 2008).

Variance components analysis—We conducted variance components analysis of the TPOCS-RS subscale scores using mixed models procedures in SAS/STAT Software 9.4 to

gauge whether targets of measurement that might impact implementation integrity influenced scores on the five TPOCS-RS subscales. Variance components analysis partitions the total variance among scores into reliable sources of variance (e.g., study group, therapist, client, time in treatment). The nested design was accounted for in the ICC calculations using mixed-model procedures (see Barber et al., 2004). Variance components were calculated using a mixed model with restricted maximum likelihood estimation for the following factors: (a) Study Group; (b) Therapist (nested within study group); (c) Client (nested within study group, therapist); (d) Time (nested within client, therapist, study group); and (e) Coder. Each factor represents a possible source of variation in treatment delivery (Barber et al., 2004). The term *study group* reflects the influence of the three groups (ICBT, YAS-ICBT, YAS-UC) on each TPOCS-RS subscale score; the term *therapist* represents systematic differences across therapists on each TPOCS-RS subscale score; the term *client* reflects systematic differences in TPOCS-RS subcale scores across each client; the term *time* reflects the effect time in treatment (measured in weeks since the intake) has on each TPOCS-RS subscale score; the term *coder* reflects systematic differences in coder ratings (tendency to score high or low) on a given TPOCS-RS subscale. 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 was run for each TPOCS-RS subscale score. We hypothesized that the majority of the variance in coders' ratings would be accounted for by study group on the subscales expected to differentiate manual-guided CBT from non-manualized UC (Barber et al., 2004): Cognitive, Behavioral, Psychodynamic, and Family.

Subscale discriminability—We examined the discriminant validity of the TPOCS-RS subscale scores by evaluating whether the subscale scores could detect expected differences between an EBT and UC. To determine if the TPOCS-RS subscale scores worked as expected, we produced adjusted least square means (LSMs) for each subscale score derived from the mixed-model analysis described in the previous section (i.e., variance components analysis). The LSMs are mean scores that are corrected for the influence of other variables. In this case, the LSMs were adjusted to account for the variables entered into the variance components analysis (i.e., study group, therapist, client, time in treatment, coder). Following an overall *F* test, the adjusted means for each subscale score were compared using pairwise comparisons with a Bonferroni adjusted alpha of .003. As *Coping Cat* is a child-focused CBT program, we hypothesized that the (a) ICBT groups (ICBT, YAS-ICBT) would have higher scores than YAS-UC on the TPOCS-RS Cognitive and Behavioral subscales, (b) YAS-UC would have higher scores than the ICBT groups on the Psychodynamic and Family subscales, and (c) the three groups would have equal scores on the Client-Centered subscale.

Results

Evaluation of Reliability

ICCs ranged from 0 to .96 (*M* = 0.76, *SD* = .18) for the individual item scores, and 0.72 to 0.94 (*M* = 0.86, *SD* = 0.12) for the subscale scores (see Table 2). The interrater reliability for 25 of the 42 items fell in the "excellent" range, 14 items fell in the "good" range, one fell in the "fair" range, and two items (*Functional Analysis*, *Behavioral Activation*) fell in the

"poor" range. The low ICCs for the *Functional Analysis* ($M = 1.07$, $SD = 0.25$; range 1 to 3.5) and *Behavioral Activation* ($M = 1.00$, $SD = 0.06$; range 1 to 2) items are likely due to limited variation ($\sigma^2 = 0.06$ and $\sigma^2 = 0.004$, respectively). Most TPOCS-RS items did not display a restricted range. Twenty-six of the items displayed the full range of scores (i.e., 1 to 7), whereas the remaining items displayed a range of at least 4.5 (i.e., 1 to 5.5). Because the *Functional Analysis* and *Behavioral Activation* items displayed low variability and poor interrater reliability, we dropped these items from subsequent analyses.

Discriminant Validity

As seen in Table 3, all of the correlations among the five TPOCS-RS subscales were small to medium in strength (all *r*s < .37) except for the correlation between the Cognitive and Behavioral subscale scores ($r = 0.59$, $p < 0.001$). To better understand this finding, we examined whether the correlation between the Cognitive and Behavioral subscale scores varied across the three groups. The correlation between the subscale scores did not differ significantly across the ICBT ($r = 0.38$, $p < 0.001$), YAS-ICBT ($r = 0.44$, $p < 0.001$), and YAS-UC ($r = 0.46$, $p < 0.001$) groups ($p > .220$). Taken together, these findings indicate that, aside from the Cognitive and Behavioral subscale scores, the TPOCS-RS subscale scores overlap to a modest degree. Furthermore, even with the relatively strong association between the Cognitive and Behavioral subscale scores, none of the subscales appear redundant.

We assessed the magnitude of the correlations between the TPOCS-RS subscale scores and scores on observational measures of the alliance (TPOCS-A) and client involvement (CIRS). As can be seen in Table 3, only the correlation between the scores on the Cognitive subscale and the TPOCS-A ($r = 0.41$) was large in magnitude. The remaining correlations fell in the small to medium range ($M r = 0.26$ of the absolute value of the correlations; $SD = 0.11$). The mean of the absolute value of the correlations between scores on the TPOCS-RS subscales and the TPOCS-A ($M r = 0.34$; $SD = 0.05$) was significantly higher than the mean of the absolute value of the correlations between scores on the TPOCS-RS subscales and the CIRS ($M r = 0.18$; $SD = 0.09$, $z = 3.62$, $p < 0.001$). The same coders scored the TPOCS-RS and TPOCS-A, so these higher correlations may be due to common-method variance. Together, these findings support the discriminant validity of the TPOCS-RS subscale scores.[2]

Variance Components Analysis

As expected, a substantial proportion of the variance in the coders' ratings of the TPOCS-RS Cognitive and Behavioral subscale scores was accounted for by study group (0.41 and 0.43, respectively). A proportion of the variance in the coders' ratings of the Psychodynamic and Family subscales were also accounted for by study group (0.14 and 0.25, respectively), but notably little variance in the Client-Centered subscale score was accounted for by study group (0.01). These findings indicate that TPOCS-RS subscale scores, except for scores on

---

[2]As an additional validity check, we assessed the magnitude of the correlations between the TPOCS-RS subscale scores and the Coping Cat Brief Adherence Scale (Kendall, 1994; Kendall et al., 1997) in the YAS-ICBT condition (N = 11 child participants) as the data from this scale was not available in the ICBT group. The magnitude of the correlations supported the convergent validity of the TPOCS-RS Cognitive ($r = 0.83$, $p < 0.01$) and Behavioral ($r = 0.64$, $p < 0.05$) subscale scores as well as the discriminant validity of the Psychodynamic ($r = -0.27$, ns), Family ($r = 0.16$, ns), and Client-Centered ($r = 0.40$, ns) subscale scores.

the Client-Centered subscale, systematically differ across the study groups. Time in treatment also accounted for a substantial proportion of the variance in scores across the TPOCS-RS subscales (ranging from 0.23 to 0.56, *Mdn* = 0.42). This finding indicates that the scores on each subscale likely varied over the course of treatment. Neither client nor coder accounted for more than 5% of the total variance in ratings; however, therapist did account for a proportion of variance in the scores on the Client-Centered and Psychodynamic subscales (0.20 and 0.16, respectively). This indicates that some therapists delivered more client-centered and psychodynamic interventions than other therapists.

Subscale Discriminability

As seen in Table 5, ICBT and YAS-ICBT had significantly higher scores than YAS-UC on the Cognitive and Behavioral subscales. Somewhat unexpectedly, ICBT had significantly higher scores than YAS-ICBT on the Cognitive and Behavioral subscales. On the Psychodynamic and Family subscales, YAS-UC had significantly higher scores than ICBT and YAS-ICBT; YAS-ICBT also had significantly higher scores than ICBT. ICBT had significantly higher scores than YAS-ICBT and YAS-UC on the Client-Centered subscale.

Discussion

The present study replicated and extended past psychometric work on a measure of treatment differentiation (TPOCS-RS) by examining the performance of the measure in a sample of youth diagnosed with a primary anxiety disorder receiving a manual-based CBT program or non-manualized UC in different settings. The psychometric findings support the use of the measure. Trained coders produced acceptable interrater reliability on the TPOCS-RS item scores and the theory-based subscale scores. Results provided evidence contributing to the construct validity of the TPOCS-RS subscale scores, demonstrating that the scores were distinct from each other *and* from independent ratings of the alliance and client involvement. Moreover, the TPOCS-RS showed promise for assessing treatment differentiation in implementation research. First, the subscale scores were sensitive to systematic variation across study groups, therapists, clients, and time. Second, the subscale scores discriminated between manual-based CBT and non-manualized UC. Together, our findings have psychometric and methodological implications.

Our results suggest that our observational measure of treatment differentiation can be coded reliably by independent coders across a manual-guided CBT program delivered in research and practice settings and non-manualized UC. Our interrater reliability statistics were comparable to those reported for other observational treatment integrity measures at the item (e.g., Barber, Mercer, Krakauer, & Calvo, 1996; Hogue et al., 2008) and subscale (e.g., Carroll et al., 2000) level. Moreover, our interrater reliability statistics were similar to those reported in earlier research with previous versions of the measure (Garland et al., 2010; McLeod & Weisz, 2010; Southam-Gerow et al., 2010; Weisz et al., 2009; Wood et al., 2006). Two of the 42 items did not display acceptable interrater reliability (i.e., *Behavioral Activation, Functional Analysis*), though the *Functional Analysis* item has displayed "excellent" reliability in a previous study (ICC = .76; McLeod & Weisz, 2010). Analyses indicated that the poor reliability for scores on the two items was likely due to low variation

in the scores. As these interventions are not found in the *Coping Cat* program, it is not surprising there was little variation in the item scores within the current sample. Because the low interrater reliability does not appear to be a result of poor item quality, we recommend that the items be retained and further evaluated in samples in which the interventions are more likely to occur (e.g., CBT for depression). Considered together, our findings indicate that highly trained coders reliably rated the TPOCS-RS items and subscales.

Our findings contribute to the evidence for the discriminant validity of the TPOCS-RS subscale scores. Correlations among the TPOCS-RS subscale scores were low to medium (Cohen, 1992), with the exception of the Cognitive and Behavioral subscale scores, which were related more strongly ($r = .59$). These correlations are consistent with past observational (Carroll et al., 2000) and self-report (Hogue, Dauber, & Henderson, 2012; Weersing et al., 2002) studies of integrity measures, supporting the notion that the TPOCS-RS subscale scores assess distinct therapy processes. As expected, associations between the most conceptually-connected subscales—the Cognitive and Behavioral subscales—was strongest (cf. McLeod & Weisz, 2010). However, there are two reasons that we do not believe the subscales should be combined. First, the magnitude of the correlation indicates that the subscale scores are not redundant (34.8% shared variance). Second, though many manual-guided CBT programs exist, some programs just include cognitive or just include behavioral interventions (e.g., Applied Behavior Analysis; Smith, Eikeseth, Sallows, & Graupner, 2009). For potential applicability in these instances there is value in retaining separate subscales.

The small to medium correlations of the TPOCS-RS subscale scores with measures of the alliance and client involvement further support the discriminant validity of the TPOCS-RS scores. The stronger correlations observed between the TPOCS-RS subscale scores and scores on the alliance measure may be due to common-method variance as the same coders scored both measures. Importantly, the magnitude of the correlations between the TPOCS-RS subscale scores and scores on the alliance measure was consistent with previous research (Carroll et al., 2000; Hogue et al., 2008).

Our third and final examination of the discriminant validity of the TPOCS-RS subscale scores was achieved by our group comparisons. As hypothesized, both ICBT groups had higher scores on the Cognitive and Behavioral subscales. The UC group also had higher scores on the Psychodynamic and Family subscales. However, compared to YAS-ICBT, the ICBT group had significantly higher scores on the Cognitive and Behavioral subscales and significantly lower scores on the Psychodynamic and Family subscales. This suggests that the YAS-ICBT group may have delivered a slightly lower dose of prescribed interventions and had more protocol violations than the ICBT group. Unexpectedly, the ICBT group had higher scores than the YAS- ICBT and YAS-UC groups on the Client-Centered subscale. Overall, these group comparison findings provide important support for the validity of the subscale scores by demonstrating that the subscale scores can discriminate between groups. Moreover, the fact that the TPOCS-RS subscale scores can detect differences between the same CBT program delivered in research and practice settings suggests that the measure is well-suited for detecting the types of subtle differences in delivery that may be present when an EBT is delivered across various settings. Of course, an important next step is to

demonstrate that the TPOCS-RS can discriminate between other forms of treatments. Although the present findings indicate that the Cognitive and Behavioral subscales can assess the successful implementation of CBT, there is less evidence that the Psychodynamic, Family, and Client-Centered subscales can assess the implementation of the types of therapy they were designed to measure. In sum, the distinctness of the subscale scores from each other and from measures of alliance and involvement along with the group comparison results all support the discriminant validity of the TPOCS-RS subscale scores and indicate that they measure distinct therapy processes.

We also found several factors that accounted for systematic variation in the TPOCS-RS subscale scores, though the factors that accounted for variation differed across the subscales. As hypothesized, study group accounted for a high proportion of non-error variance in scores on the four subscales that the ICBT and UC groups were expected to differ on: Cognitive, Behavioral, Psychodynamic, and Family. Interestingly, time in treatment also accounted for a high proportion of the non-error variance in all subscale scores, consistent with previous research (Barber et al., 2004; Hogue et al., 2008). If integrity scores do vary from session to session, as these findings suggest, then producing integrity scores by averaging across sessions may obscure important information (Boswell et al., 2013). Therapist accounted for a higher proportion of non-error variance than did study group in the Client-Centered and Psychodynamic subscale scores, indicating that use of client-centered and psychodynamic interventions was more related to therapist than study group. The relatively low proportion of non-error variance explained by therapist in the Cognitive and Behavioral subscale scores is probably explained by the fact that therapists in the ICBT groups were trained and supervised on these procedures, which likely served to reduce variability across therapists. Previous research has found that the proportion of variance explained by the therapist does vary across scores on scales designed to capture different therapeutic approaches (see Hill, O'Grady, & Elkin, 1992; Hogue et al., 2008). In contrast, client effects were not observed, suggesting that the delivery of interventions did not vary across clients. However, as there were not many clients nested in therapists, our ability to tease out client and therapist effects may be limited. An important direction for future research is to identify whether integrity scores vary at the therapist or client level so researchers can isolate ways to maintain integrity in various settings (Boswell et al., 2013). In all, these findings suggest that the TPOCS-RS subscale scores may be sensitive to variations due to therapist-, time-, and group-level factors. Thus, the TPOCS-RS may aid efforts to identify systematic sources of variation in implementation integrity scores.

Considered together, our findings have important implications for how scores on the TPOCS-RS subscales might be used and interpreted in future research. Scores on the Cognitive and Behavioral subscales significantly differed across ICBT and YAS-ICBT groups, suggesting that the implementation of ICBT differed across settings. That the implementation of ICBT appears to differ across research and practice settings has implications for future implementation research. It may be possible to use the TPOCS-RS to identify minimal integrity scores to serve as a "benchmark" for determining whether a session or course of treatment was delivered consistent with the intended approach (Addis, 1997; Sholomskas et al., 2005; Shaw, 1984). The benchmarking method is a promising tool for advancing knowledge about the integrity of EBT implementation (McLeod et al., 2013).

Benchmarking studies could be designed to evaluate whether therapist performance in community settings approximate the performance standards achieved by therapists in efficacy trials. To date, the benchmarking method has primarily been used to identify treatment outcome targets and not therapist performance targets (e.g., Merrill, Tolbert, & Wade, 2003; Wade, Treat, & Stuart, 1998). However, benchmarking methods could also be used to identify therapist performance targets for adherence, differentiation, and/or competence. Such data would speak directly to the question of the degree to which practitioners are demonstrating treatment integrity.

Our findings have other implications for future research. A primary aim of this study was to expand the coverage of the TPOCS-RS to facilitate research examining a broader variety of treatment approaches. The measure now includes a longer list of interventions found in EBTs and used by therapists in UC. Further, the new items and resulting subscale scores possess solid psychometric properties, offering one alternative in response to calls for the development of psychometrically strong measures to assess implementation integrity (Garland et al., 2010; Schoenwald et al., 2011). Thus, the TPOCS-RS represents a strong candidate as one measure in future implementation research endeavors. For example, as noted above, therapist training studies could leverage the TPOCS-RS to set training benchmarks, determine the effectiveness of the training, and/or compare multiple training approaches. Another important future direction concerns the understanding of how treatment is implemented in practice settings, including the increasing number of EBT implementation studies. The TPOCS-RS is well-suited to be part of such studies and can be used to determine whether (and how) therapists' implementation of EBTs in practice settings may deviate from the treatment manual and whether this influences treatment outcomes.

Despite its strengths, the TPOCS-RS is a resource-intensive measure. It takes up to three months to train coders to reliability and the scoring of each session takes approximately one hour. Thus, another important future research direction concerns using the TPOCS-RS (or a similar observational integrity measure) as a criterion measure to develop cost-effective, therapist-report measures designed to assess implementation integrity (Hogue et al., 2012).

Although the findings illustrate the research potential of the TPOCS-RS, a few limitations of the study should be considered. First, no sample of therapists and settings can be completely representative of how manual-guided, or non-manualized, treatment is delivered. Thus, it is possible that a different pattern of findings would emerge in a different sample. Second, our ability to model therapist effects in the variance components analysis was restricted by the fact that there was a limited number of clients nested within therapists, particularly in the YAS sample. Third, we were not able to code all sessions from each case. It therefore is possible that the coded sessions may not fully represent the therapy provided in these three groups.

Although the present findings support the validity of the Cognitive and Behavioral subscale scores, the validity evidence for scores on the Psychodynamic, Family, and Client- Centered subscales remains relatively limited, suggesting a direction for future research. In such research, it will be important to assess the validity of the Psychodynamic, Family, and Client-Centered subscale scores by applying the TPOCS-RS to therapies known by some

independent means (e.g., manualized forms of family or psychodynamic therapy) to faithfully represent each of these therapeutic approaches.

Certain strengths of the study also warrant attention. We coded sessions from a large sample of youth diagnosed with primary anxiety disorders. Therapists delivered the same EBT, *Coping Cat*, across research and practice settings. We coded all available sessions from each case, thus addressing critiques that past work has not coded a sufficient number of recordings to accurately characterize treatment delivery (e.g., Barber et al., 2004). Finally, we used psychometrically sound alliance and client involvement measures for our validity analyses.

The psychometric findings presented here suggest two broad conclusions. First, the TPOCS-RS has procedural and psychometric characteristics that give it the potential to address a methodological gap in implementation research. And second, because the TPOCS-RS assesses a wide range of interventions, it provides a relatively comprehensive description of treatment delivery and is well-suited to assessing treatment differentiation. Thus, the TPOCS-RS appears to have characteristics and strengths that give it potential to fill a measurement gap.

Acknowledgments

References

Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors. Administration and Policy in Mental Health and Mental Health Services Research. 2010; 38:4–23.10.1007/s10488-010-0327-7 [PubMed: 21197565]

Addis ME. Evaluating the treatment manual as a means of disseminating empirically validated psychotherapies. Clinical Psychology: Science and Practice. 1997; 4:1–11.10.1111/j. 1468-2850.1997.tb00094.x

Allen, JA.; Linnan, LA.; Emmons, KM. Fidelity and its relationship to implementation effectiveness, adaptation, and dissemination. In: Brownson, RC.; Colditz, GA.; Proctor, EK., editors. Dissemination and implementation research in health: translating science to practice. New York, NY: Oxford University Press; 2012. p. 281-304.

Barber JP, Foltz C, Crits-Christoph P, Chittams J. Therapists' adherence and competence and treatment discrimination in the NIDA collaborative cocaine treatment study. Journal of Clinical Psychology. 2004; 60:29–41.10.1002/jclp.10186 [PubMed: 14692007]

Barber JP, Mercer D, Krakauer I, Calvo N. Development of an adherence/competence rating scale for individual drug counseling. Drug and Alcohol Dependence. 1996; 43:125–132.10.1016/S0376-8716(96)01305-1 [PubMed: 9023068]

Bearman, SK.; Weisz, JR.; McLeod, BD. Unpublished scoring manual. Judge Baker Children's Center, Harvard University; Boston, MA: 2010. Scoring manual for the Therapy Process Observational Coding System for Child Psychotherapy – Group Psychotherapy Scale.

Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, Czajkowski S. Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH behavior change consortium. Health Psychology. 2004; 23:443–451.10.1037/0278-6133.23.5.443 [PubMed: 15367063]

Boswell JF, Gallagher MW, Sauer-Zavala SE, Bullis J, Gorman JM, Shear MK, Barlow DH. Patient characteristics and variability in adherence and competence in cognitive-behavioral therapy for

panic disorder. Journal of Consulting and Clinical Psychology. 2013; 81:443–454.10.1037/a0031437 [PubMed: 23339537]

Carroll KM, Nich C, Sifty RL, Nuro KF, Frankfurter TL, Ball SA, Rounsaville BJ. A general system for evaluating therapist adherence and competence in psychotherapy research. Drug and Alcohol Dependence. 2000; 57:225–238.10.1016/S0376-8716(99)00049-6 [PubMed: 10661673]

Center for the Study and Prevention of Violence. Blueprints for violence prevention. 2010. Retrieved June 3, 2013, from http://www.colorado.edu/cspv/blueprints/

Chu BC, Kendall PC. Positive association of child involvement and treatment outcome within a manual-based cognitive-behavioral treatment for children with anxiety. Journal of Consulting and Clinical Psychology. 2004; 72:821–829.10.1037/0022-006X.72.5.821 [PubMed: 15482040]

Chu BC, Kendall PC. Therapist responsiveness to child engagement: Flexibility within manual-based CBT for anxious youth. Journal of Clinical Psychology. 2009; 65:736–754.10.1002/jclp.20582 [PubMed: 19391153]

Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment. 1994; 6:284–290.10.1037/1040-3590.6.4.284

Cohen J. A power primer. Psychological Bulletin. 1992; 112:155–159.10.1037/0033-2909.112.1.155 [PubMed: 19565683]

Fjermestad K, McLeod BD, Heiervang ER, Havik OE, Ost LG, Haugland BSM. Factor structure and psychometric properties of the Therapy Process Observational Coding System for Child Psychotherapy Alliance scale. Journal of Clinical Child and Adolescent Psychology. 2012; 41:1–9.10.1080/15374416.2012.651999 [PubMed: 22233240]

Garland AF, Brookman-Frazee L, Hurlburt MS, Accurso EC, Zoffness RJ, Haine-Schlagel R, Ganger W. Mental health care for children with disruptive behavior problems: A view inside therapists' offices. Psychiatric Services. 2010; 61:788–795.10.1176/appi.ps.61.8.788 [PubMed: 20675837]

Garland, AF.; Brookman-Frazee, L.; McLeod, BD. Unpublished scoring manual. Child and Adolescent Services Research Center; San Diego, California: 2008. Scoring manual for the PRAC Study Therapy Process Observational Coding System for Child Psychotherapy: Strategies Scale.

Hagermoser Sanetti LM, Kratochwill TR. Toward developing a science of treatment integrity: Introduction to the special series. School Psychology Review. 2009; 38(4):445–459.

Hill CE, O'Grady KE, Elkin I. Applying the Collaborative Study Psychotherapy Rating Scale to therapist adherence in cognitive-behavior therapy, interpersonal therapy, and clinical management. Journal of Consulting and Clinical Psychology. 1992; 60:73–79. doi:0022-006X/92/J3.00. [PubMed: 1556289]

Hogue A, Dauber S, Henderson CE. Therapist self-report of evidence-based practices in usual care for adolescent behavior problems: Factor and construct validity. Administration and Policy in Mental Health and Mental Health Services Research. 2012; 41:126–139.10.1007/s10488-012-0442-8 [PubMed: 23124275]

Hogue A, Henderson CE, Dauber S, Barajas PC, Fried A, Liddle HA. Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. Journal of Consulting and Clinical Psychology. 2008; 76:544–555.10.1037/0022-006X.60.1.73 [PubMed: 18665684]

Hogue A, Liddle HD, Rowe C. Treatment adherence process research in family therapy: A rationale and some practical guidelines. Psychotherapy. 1996; 33:332–345.10.1037/0033-3204.33.2.332

Hurlburt MS, Garland AF, Nguyen K, Brookman-Frazee L. Child and family therapy process: Concordance of therapist and observational perspectives. Administration and Policy in Mental Health and Mental Health Services Research. 2010; 37:230–244.10.1007/s10488-009-0251-x [PubMed: 19902347]

Institute of Medicine. Crossing the quality chasm: A new health system for the 21st century. Washington, DC: National Academy Press; 2001.

Kendall PC. Treating anxiety disorders in children: Results of a randomized clinical trial. Journal of Consulting and Clinical Psychology. 1994; 62:100–110.10.1037/0022-006X.62.1.100 [PubMed: 8034812]

Kendall PC, Flannery-Schroeder E, Panichelli-Mindel SM, Southam-Gerow M, Henin A, Warman M. Therapy for youths with anxiety disorders: A second randomized clinical trial. Journal of Consulting and Clinical Psychology. 1997; 65:366–380.10.1037/0022-006X.65.3.366 [PubMed: 9170760]

Kendall, PC.; Hedtke, K. Cognitive-behavioral therapy for anxious children: Therapist manual. 3. Ardmore, PA: Workbook Publishing; 2006a.

Kendall, PC.; Hedtke, K. Coping cat workbook. 2. Ardmore, PA: Workbook Publishing; 2006b.

Kendall PC, Hudson JL, Gosch E, Flannery-Schroeder E, Suveg C. Cognitive-behavioral therapy for anxiety disordered youth: A randomized clinical trial evaluating child and family modalities. Journal of Consulting and Clinical Psychology. 2008; 76:282–297.10.1080/10508420802064309 [PubMed: 18377124]

Liber JM, McLeod BD, Van Widenfelt BM, Goedhart AW, van der Leeden AJM, Utens EMWJ, Treffers PDA. Examining the relation between the therapeutic alliance, treatment adherence, and outcome of cognitive behavioral treatment for children with anxiety disorders. Behavior Therapy. 2010; 41:172–186.10.1016/j.beth32009.02.003 [PubMed: 20412883]

Lichstein KL, Riedel BW, Grieve R. Fair tests of clinical trials: A treatment implementation model. Advances in Behavior Research and Therapy. 1994; 16:1–29.10.1016/0146-6402(94)90001-9

McLeod, BD. Unpublished scoring manual. University of California; Los Angeles: 2001. Therapy Process Observational Coding System for Child Psychotherapy – Strategies Scale.

McLeod, BD. Unpublished scoring manual. Virginia Commonwealth University; Richmond, VA: 2010. Therapy Process Observational Coding System for Child Psychotherapy – Revised Strategies Scale.

McLeod BD, Southam-Gerow MA, Tully CB, Rodriguez A, Smith MM. Making a case for treatment integrity as a psychosocial treatment quality indicator for youth mental health care. Clinical Psychology: Science and Practice. 2013; 20:14–32. doi:10-1111/cpsp.12020. [PubMed: 23935254]

McLeod BD, Weisz JR. The Therapy Process Observational Coding System-Alliance scale: Measure characteristics and prediction of outcome in usual clinical practice. Journal of Consulting and Clinical Psychology. 2005; 73:323–333.10.1037/0022-006X.73.2.323 [PubMed: 15796640]

McLeod BD, Weisz JR. The Therapy Process Observational Coding System for Child Psychotherapy Strategies Scale. Journal of Clinical Child & Adolescent Psychology. 2010; 39:436–443.10.1080/15374411003691750 [PubMed: 20419583]

Merrill KA, Tolbert VE, Wade WA. Effectiveness of cognitive therapy for depression in a community mental health center: A benchmarking study. Journal of Consulting and Clinical Psychology. 2003; 71:404–409.10.1037/0022-006X.71.2.404 [PubMed: 12699035]

Orlinsky, DE.; Rønnestad, MH.; Willutzki, U. Fifty years of psychotherapy process-outcome research: Continuity and change. In: Lambert, M., editor. Bergin and Garfield's handbook of psychotherapy and behavior change. 5. New York, NY: Wiley; 2004. p. 307-389.

Perepletchikova F, Kazdin AE. Treatment integrity and therapeutic change: Issues and research recommendations. Clinical Psychology: Science and Practice. 2005; 12:365–383.10.1093/clipsy/bpi045

Perepletchikova F, Treat TA, Kazdin AE. Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. Journal of Consulting and Clinical Psychology. 2007; 75:829–841.10.1037/0022-006X.75.6.829 [PubMed: 18085901]

Schoenwald SK, Garland AF, Chapman JE, Frazier SL, Sheidow AJ, Southam-Gerow MA. Toward the effective and efficient measurement of implementation fidelity. Administration and Policy in Mental Health and Mental Health Services Research. 2011; 38:32–43.10.1007/s10488-010-0321-0 [PubMed: 20957425]

Shaw, BF. Specification of the training and evaluation of cognitive therapists for outcome studies. In: Williams, JBW.; Spitzer, RL., editors. Psychotherapy research: Where are we and where should we go?. New York: Guilford Press; 1984. p. 173-188.

Sholomskas DE, Syracuse-Siewert G, Rounsaville BJ, Ball SA, Nuro KF, Carroll KM. We don't train in vain: A dissemination trial of three strategies of training clinicians in cognitive-behavioral therapy. Journal of Consulting and Clinical Psychology. 2005; 73:106–115.10.1037/0022-006X.73.1.106 [PubMed: 15709837]

Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin. 1979; 86:420–428.10.1037/0033-2909.86.2.420 [PubMed: 18839484]

Smith T, Eikeseth S, Sallows GO, Graupner TD. Efficacy of applied behavior analysis in autism. Journal of Pediatrics. 2009; 155:151–152.10.1016/j.jpeds.2009.01.065 [PubMed: 19559310]

Southam-Gerow MA, Weisz JR, Chu BC, McLeod BD, Gordis EB, Connor-Smith JK. Does cognitive behavioral therapy for youth anxiety outperform usual care in community clinics? An initial effectiveness test. Journal of the American Academy of Child and Adolescent Psychiatry. 2010; 49:1043–52.10.1016/j.jaac.2010.06.009 [PubMed: 20855049]

Wade WA, Treat TA, Stuart GL. Transporting an empirically supported treatment for panic disorder to a service clinic setting: A benchmarking strategy. Journal of Consulting and Clinical Psychology. 1998; 66:231–239.10.1037/0022-006X.66.2.231 [PubMed: 9583326]

Waltz J, Addis ME, Koerner K, Jacobson NS. Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. Journal of Consulting and Clinical Psychology. 1993; 61:620–630.10.1037/0022-006X.61.4.620 [PubMed: 8370857]

Weersing VR, Weisz JR, Donenberg GR. Development of the Therapy Procedures Checklist: A therapist-report measure of technique use in child and adolescent treatment. Journal of Clinical Child Psychology. 2002; 31:168–180.10.1207/153744202753604458

Weisz JR, Ng MY, Bearman SK. Odd couple? Re-envisioning the relation between science and practice in the dissemination-implementation era. Clinical Psychological Science. 2014; 2(1):58–74.10.1177/2167702613501307

Weisz JR, Southam-Gerow MA, Gordis EB, Connor-Smith JK, Chu BC, Langer DA, Weiss B. Cognitive-behavioral therapy versus usual clinical care for youth depression: An initial test of transportability to community clinics and clinicians. Journal of Consulting and Clinical Psychology. 2009; 77:383–96.10.1037/a0013877 [PubMed: 19485581]

Wood JJ, Piacentini JC, Southam-Gerow M, Chu BC, Sigman M. Family cognitive behavioral therapy for child anxiety disorders. Journal of the American Academy of Child and Adolescent Psychiatry. 2006; 45:314–321.10.1097/01.chi.0000196425.88341.b0 [PubMed: 16540816]

Table 1

Client Descriptive Data and Comparisons Across Groups

Variable	M (SD) or %			F or Chi Square
	ICBT	**YAS-ICBT**	**YAS-UC**	
Age	10.36 (1.90)	11.32 (2.32)	10.44 (1.91)	1.56
Gender				
Male	60.80	29.40	52.40	5.04
Ethnicity				29.91[***]
Caucasian	86.30[a]	41.20	33.30	
African-American	9.80	-	9.50	
Latino	2.00	17.60[d]	42.90[e]	
Mixed/Other	2.00	5.90	9.50	
Not Reported	-	35.30[b]	4.80	
CBCL				
Total	63.18 (8.44)	64.19 (7.34)	65.06 (6.46)	0.39
Internalizing	67.40 (8.37)	66.38 (8.33)	66.82 (8.33)	0.10
Externalizing	52.96 (10.08)	60.81 (7.49)[d]	59.41 (9.67)	5.61[**]
Primary Diagnoses				22.73[**]
GAD	37.30[c]	5.90	14.30	
SAD	29.40	35.30	38.10	
SOP	33.30	23.50	28.60	
SP	-	35.30[b]	19.00	
Family Income				15.66[***]
Up to 60k per year	35.30	70.60[d]	76.20[e]	
Number of Sessions	15.92 (1.43)	16.82 (5.02)	15.71 (9.34)	0.26
Weeks in Treatment	19.52 (3.97)	26.38 (10.41)[d]	26.84 (15.53)[e]	6.45[**]

Note. ICBT = individual cognitive-behavioral therapy delivered in Kendall et al. study, YAS-ICBT = ICBT delivered in YAS, YAS-UC = usual care delivered in YAS. For continuous variables, an ANOVA was conducted. For categorical variables, chi square analyses were conducted. CBCL = Child Behavior Checklist, GAD = generalized anxiety disorder, SAD = separation anxiety disorder, SOP = social phobia, SP = specific phobia.

[a]ICBT > YAS-ICBT, YAS-UC.

[b]YAS-ICBT > YAS-UC.

[c]ICBT > YAS-ICBT.

[d]YAS-ICBT > ICBT.

[e]YAS-UC > ICBT.

[*]$p < 0.05$,

[**]$p < 0.01$,

[***]$p < 0.001$

Table 2

TPOCS-RS Item Descriptions and Interrater Reliability

TPOCS-RS Item Description	ICC
Cognitive Subscale	0.94
General Cognitive Focus: Extent to which a therapist uses cognitive interventions in a session.	0.92
Cognitive Education: Teaches cognitive model/Identifies how the cognitive model applies to an aspect of the client's life.	0.88
Cognitive Distortion: Teaches/encourages client to identify/restructure cognitive distortions.	0.93
Coping Skills: Teaches/encourages client to use coping skills (e.g., problem-solving skills).	0.90
Behavioral Subscale	0.94
General Behavioral Focus: Extent to which the therapist employs behavioral interventions.	0.92
Functional Analysis of Behavior: Performs functional analysis/teaches A-B-C model.	0.29
Relaxation Interventions: Teaches/encourages client to use relaxation.	0.96
Respondent Interventions: Develops a fear hierarchy and/or performs an exposure.	0.95
Operant Strategies: Teaches principles of operant interventions, sets up operant system, or employs operant intervention.	0.84
Skill Building: Interventions that focus on building behavioral skills (e.g., social skills).	0.86
Behavioral Activation: Teaches/demonstrates relation between pleasant activities and mood improvement or assigns participation in a pleasant event to improve mood.	0.00
Monitoring: The therapist elicits measurements of a client's mood, functioning, or experience, or encourages a client to monitor his/her mood.	0.87
Modeling: Teaches specific skills using observational learning methods.	0.80
Psychodynamic Subscale	0.74
Psychodynamic Focus: Extent to which the therapist employs psychodynamic interventions.	0.76
Addresses Transference: Discusses or interprets the client's interaction with the therapist.	0.71
Explores Past: Discusses client's past experiences.	0.63
Addressing Client Resistance: Identifies/processes client's resistance to therapy and/or resistance to change.	0.63
Interpretation: Comments on client behavior and/or relates that behavior to an aspect of the client's characteristics, general functioning, and/or past experiences.	0.73
Family Subscale	0.94
General Family Focus: Extent to which the therapist employs family interventions in a session.	0.94
Targets Other Participants: Participants other than the target child are asked to modify their affect, behavior, cognitions.	0.95
Recruits Others: Tries to recruit/retain parents and other family members for future sessions.	0.54
Parenting Style: Helps and/or encourages parents to modify their parenting practices.	0.71
Multiparticipant Interactions: Establishes/teaches/discusses in-session interactions.	0.87
Family Members' Roles: Therapist teaches or emphasizes how problems may be caused/maintained by family dynamics.	0.82

TPOCS-RS Item Description	ICC
Client-Centered Subscale	
General Client-Centered Focus: Extent to which the therapist validates client's feelings or elicits client's perspective.	0.72
Validates Client: Validates client's feelings and/or treatment goals.	0.62
Positive Regard: Responds to client in warm and compassionate manner.	0.64
Client Perspective: Attempts to understand client's point of view/Probes for client's unique perspective.	0.77
	0.63
General Items	
Rehearsal: Encourages client to participate in hypothetical enactments.	0.86
Homework: Assigns and/or reviews homework assignments.	0.87
Play/Art Therapy: Therapist utilizes play or art as a form of therapy.	0.83
Encourages Affect: Discusses affect and/or encourages client to express affect.	0.72
Session Goals: Establishes/reviews session goals.	0.66
Treatment Goals: Establish treatment goals/Encourage client to discuss treatment goals.	0.82
Previous Themes: Comments on themes from previous sessions/builds on past successes.	0.60
Psychoeducation: Teaches client about general psychological principles (e.g., anxiety).	0.84
Questioning: Therapist asks about general topics.	0.72
Self-Disclosure: Therapist provides information about his/her personal life, feelings, experiences.	0.80
Advice: Therapist provides the client with precise instructions on how to address a specific issue.	0.68
Coaching: Therapist makes efforts to actively direct the client as they practice a specific skill.	0.88
Assessment: Therapist gathers data about psychiatric, social, academic, or medical problems.	0.69

Note. TPOCS-RS = The Therapy Process Observational Coding System for Child Psychotherapy – Revised Strategies Scale; ICC = intraclass correlation coefficient.

Table 3

Correlations among TPOCS-RS Subscales, Alliance, and Involvement

	1.	2.	3.	4.	5.	6.	7.
1. Cognitive	**0.89**	0.59***	−0.32***	−0.43***	0.34***	0.41***	0.22***
2. Behavioral		**0.79**	−0.29***	−0.33***	0.29***	0.31***	0.07**
3. Psychodynamic			**0.79**	0.19***	0.02*	−0.43***	−0.27***
4. Family				**0.88**	−0.04	−0.29***	−0.23***
5. Client-centered					**0.59**	0.26***	0.11**
6. Alliance						**0.81**	0.76***
7. Involvement							**0.85**

Note: TPOCS-RS = The Therapy Process Observational Coding System for Child Psychotherapy – Revised Strategies scale. The numbers in bold represent the internal consistency of the TPOCS-RS subscales.

*
p < 0.05,

**
p < 0.01,

***
p < 0.001.

Table 4

Variance Components for TPOCS-RS Subscales

TPOCS-RS Subscale	Variance Components					
	Group	Therapist	Client	Time	Coder	Residual
Cognitive subscale	0.41	0.03	0.05	0.42	<0.01	0.10
Behavioral subscale	0.43	0.02	0.03	0.43	<0.01	0.09
Psychodynamic subscale	0.14	0.16	0.05	0.23	0.04	0.37
Family subscale	0.25	0.08	<0.01	0.56	<0.01	0.10
Client-centered subscale	0.01	0.20	0.03	0.31	<0.01	0.44

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 5

Least Square Means of TPOCS-RS Subscales Across Groups

Subscales	M			F
	ICBT	**YAS-ICBT**	**YAS-UC**	
Cognitive	3.79[a]	3.08[b]	1.43	136.75[***]
Behavioral	2.66[a]	1.80[b]	1.20	63.22[***]
Psychodynamic	1.08	1.24[c]	1.42[d]	16.60[***]
Family	1.34	1.51[c]	2.22[d]	96.97[***]
Client-centered	2.88[a]	2.55	2.67	4.11[*]

Note. TPOCS-RS = The Therapy Process Observational Coding System for Child Psychotherapy – Revised Strategies scale. ICBT = individual cognitive-behavioral therapy delivered in Kendall et al. study, YAS-ICBT = ICBT delivered in YAS, YAS-UC = usual care delivered in YAS. Pairwise comparisons were conducted with a Bonferonni adjusted alpha; superscripts represent significance at $p < .003$.

[a]ICBT > YAS-ICBT, YAS-UC.

[b]YAS-ICBT > YAS-UC.

[c]YAS-ICBT > ICBT.

[d]YAS-UC > ICBT, YAS-ICBT.

[*] $p < 0.05$,

[**] $p < 0.01$,

[***] $p < 0.001$.