

Does the Delivery of CBT for Youth Anxiety Differ Across Research and Practice Settings?

Meghan M. Smith

Bryce D. McLeod

Michael A. Southam-Gerow

Virginia Commonwealth University

Amanda Jensen-Doss

University of Miami

Philip C. Kendall

Temple University

John R. Weisz

Harvard University

Does delivery of the same manual-based individual cognitive-behavioral treatment (ICBT) program for youth anxiety differ across research and practice settings? We examined this question in a sample of 89 youths (M age = 10.56, SD = 1.99; 63.70% Caucasian; 52.80% male) diagnosed with a primary anxiety disorder. The youths received (a) ICBT in a research setting, (b) ICBT in practice settings, or (c) non-manual-based usual care (UC) in practice settings. Treatment delivery was assessed using four theory-based subscales (Cognitive-behavioral, Psychodynamic, Client-Centered, Family) from the Therapy Process Observational Coding System for Child Psychotherapy–Revised Strategies scale (TPOCS-RS). Reliable independent coders, using the TPOCS-RS, rated 954 treatment sessions from two randomized controlled trials (1 efficacy and 1

effectiveness trial). In both settings, therapists trained and supervised in ICBT delivered comparable levels of cognitive-behavioral interventions at the beginning of treatment. However, therapists trained in ICBT in the research setting increased their use of cognitive-behavioral interventions as treatment progressed whereas their practice setting counterparts waned over time. Relative to the two ICBT groups, the UC therapists delivered a significantly higher dose of psychodynamic and family interventions and a significantly lower dose of cognitive-behavioral interventions. Overall, results indicate that there were more similarities than differences in manual-based ICBT delivery across research and practice settings. Future research should explore why the delivery of cognitive-behavioral interventions in the ICBT program changed over time and across settings, and whether the answers to these questions could inform implementation of ICBT programs.

Preparation of this article was supported by a grant from the National Institute of Mental Health (RO1 MH86529; McLeod & Southam-Gerow).

Address correspondence to Meghan M. Smith, Department of Psychology, Virginia Commonwealth University, 806 West Franklin Street, PO Box 842018, Richmond, VA 23284-2018; e-mail: smithmm9@vcu.edu.

Keywords: Youth; Cognitive-Behavioral Treatment; Implementation; Evidence-Based

GIVEN THE DOCUMENTED DIFFERENCES between the clients, therapists, and contexts in research and practice settings (e.g., Southam-Gerow, Rodriguez,

Chorpita, & Daleiden, 2012), it is possible that the delivery of evidence-based treatments (EBTs) may vary across settings (Weisz et al., 2013). Practice and research settings differ across several important dimensions (e.g., youth, therapist, and setting characteristics; Southam-Gerow, Weisz, & Kendall, 2003) that may influence treatment delivery (i.e., those aspects of treatment under the influence of the therapist). It is possible that one, or more, youth, therapist, or setting factors could influence treatment delivery in practice settings, the place where most youth receive mental health care services (Schoenwald & Hoagwood, 2001; Schoenwald et al., 2011). Since treatment delivery is linked with clinical outcomes (Hogue et al., 2008), it is important to understand if differences in treatment delivery across settings do exist. For this reason, treatment delivery is considered an important outcome domain in implementation research, a field focused on the process and outcome of transporting EBTs from research to practice settings (Proctor et al., 2011).

Treatment integrity represents an important measurement domain that can be used to characterize variations in treatment delivery across settings. Treatment integrity is composed of several components, including *treatment adherence*, *treatment differentiation*, and *therapist competence* (Bellg et al., 2004; McLeod, Southam-Gerow, Tully, Rodriguez, & Smith, 2013; Perepletchikova & Kazdin, 2005; Waltz, Addis, Koerner, & Jacobson, 1993). Two of these components, treatment adherence and differentiation, focus on the type and quantity (i.e., dosage) of the interventions being delivered, with adherence reflecting the extent to which interventions prescribed by the treatment model were delivered and differentiation capturing the extent to which a wider range of interventions were delivered. Though most efforts to characterize delivery in the treatment literature have thus far focused on adherence measurement (Perepletchikova, Treat & Kazdin, 2007), some have questioned whether this is sufficient when trying to discern how setting or context may influence treatment delivery (Garland et al., 2010; Southam-Gerow & McLeod, 2013).

Treatment differentiation may be an important component to assess when trying to assess differences in delivery across settings (McLeod et al., 2013). First, most differentiation instruments assess interventions prescribed by a treatment and thus provide an index of treatment adherence, though less precise than an instrument designed solely for adherence. Second, differentiation permits a rich description of the delivery of other observed interventions, including interventions that are proscribed by the treatment model (e.g., in the case of cognitive-behavioral

treatment [CBT] for youth anxiety, psychodynamic interventions) and those that are permitted but not explicitly prescribed by the treatment model (e.g., in the case of CBT for youth anxiety, client-centered interventions).

When EBTs are transported to practice settings, several factors increase the likelihood that interventions (proscribed or otherwise) not found in the treatment model may be delivered (Garland et al., 2010; McLeod et al., 2013). Therapists in practice settings have varied training backgrounds and experience (Santa Ana et al., 2009), and some have hypothesized that these factors may increase the likelihood these therapists deliver a range of interventions, including some that are proscribed by the treatment model, when delivering an EBT (e.g., Southam-Gerow et al., 2010; Weisz et al., 2009). In contrast, therapists in research settings tend to have more specialized training backgrounds (Bearman et al., 2013; Herschell, Kolko, Baumann, & Davis, 2010), which may help decrease the likelihood that they will deliver proscribed interventions. Regarding youth factors, the youth seen in practice settings are more demographically and clinically heterogeneous (Southam-Gerow et al., 2003; Southam-Gerow, Marder, & Austin, 2008); it is thus possible that therapists make adaptations to EBTs based on those differences with the hope of making them effective in practice settings (Weisz, Jensen, & McLeod, 2005). Adherence measurement would exclude information about these other interventions (McLeod et al., 2013). Thus, when attempting to characterize treatment delivery in practice settings, it may be important to use a differentiation instrument that captures treatment content beyond adherence to one specific protocol.

Though it is reasonable to assume that EBT delivery may vary across settings, to our knowledge this assumption has not been tested. Research that directly evaluates this assumption could be used to ensure that EBTs delivered in practice settings are implemented along lines consistent with the treatment model (Bond, Becker, & Drake, 2011; Schoenwald, 2011). For example, researchers could use data on differentiation to characterize the variability (or lack thereof) of delivery of an EBT by therapists in practice settings compared to the EBT when delivered in research settings (McLeod et al., 2013).

In this study, we evaluated whether the delivery of an efficacious individual CBT (ICBT; i.e., *Coping Cat*; Kendall & Hedtke, 2006) program for youth diagnosed with anxiety disorders varies across research and practice settings. The *Coping Cat* program is well suited for testing this assumption as it was developed and evaluated in a research setting (Kendall, Hudson, Gosch, Flannery-Schroeder, &

Suveg, 2008) and then subsequently tested across diverse practice settings (Southam-Gerow et al., 2010). Discrepancies in training techniques can contribute to differences in treatment delivery across settings (Henggeler, Schoenwald, Liao, Letourneau, & Edwards, 2002; Sholomskas et al., 2005). We thus selected studies conducted in research and practice settings that used the same “gold-standard” procedures to train the therapists in ICBT (i.e., training workshop, treatment manual, and supervision; Sholomskas et al., 2005), thereby standardizing training procedures across settings. We also included the usual care (UC) control group from the practice settings. In Southam-Gerow et al. (2010), therapists volunteered to participate and were randomly assigned to receive training in ICBT or to a “usual care” control group. In the present study, the UC group also represents a control group, allowing us to provide perspective on variations in treatment delivery related to setting alone versus setting plus training.

We also assessed whether treatment delivery changed over time, consistent with studies of other treatments (Boswell et al., 2013; Henggeler, Sheidow, Cunningham, Donohue, & Ford, 2008). For example, some researchers have found that youth- and caregiver-report of treatment adherence increased over treatment (Henggeler et al., 2008), whereas others have found treatment adherence and competence decreased over treatment (Boswell et al., 2013). We thus sought to learn whether EBT delivery might change over time and whether patterns of change differ in practice versus research settings.

We tested two main hypotheses. Research suggests that the dose of CBT delivered by therapists in research settings may be high (Wood, Piacentini, Southam-Gerow, Chu, & Sigman, 2006), especially compared to the dose of CBT delivered by therapists in practice settings (Weisz et al., 2009). Thus, we hypothesized that therapists delivering ICBT in the research setting (in comparison to ICBT in the practice settings and UC) would deliver the highest dose of interventions that align with the ICBT protocol (i.e., interventions consistent with Coping Cat and client-centered domains), and that therapists delivering ICBT in the practice settings would deliver higher doses of interventions that align with the ICBT protocol than therapists delivering UC in the practice settings. Research indicates that therapists trained to deliver CBT for youth depression in practice settings delivered nonprotocol interventions (Weisz et al., 2009). We thus hypothesized that therapists delivering UC (in comparison to the other two ICBT groups) would deliver the highest dose of interventions not found in the protocol (i.e., interventions consistent

with the Family and Psychodynamic domains), and that therapists delivering ICBT in the practice settings would deliver higher doses of interventions not found in the protocol than therapists delivering ICBT in the research setting. Given the contradictory findings from prior literature regarding changes in treatment delivery over time (Boswell et al., 2013; Henggeler et al., 2008), we did not formulate specific hypotheses about change over time.

Method

PARTICIPANTS AND STUDY SITES

Participants for the study included 89 youth participants from two randomized controlled trials conducted by separate research groups. Recorded treatment sessions (both audio and video) were analyzed for this study. Inclusion criteria for these recordings included: (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 youth were aged 7–15 years (M age = 10.56, SD = 2.00; 65.20% Caucasian; 52.80% male) and met diagnostic criteria for a primary anxiety disorder (see Table 1).

Research Setting

Kendall et al. (2008) compared the efficacy of ICBT, family-CBT, and an active control group. Only the ICBT group was used in this study. Fifty-one youth participants (M age = 10.36, SD = 1.90; 86.20% Caucasian; 60.80% male) received ICBT at a research clinic that specialized in the treatment of anxiety disorders. Therapists (n = 16; 12.50% male) were mostly Caucasian (81.25%); some were Latino and Asian/Pacific Islander (both 6.25%), and 6.25% did not report their ethnicity/racial background. These therapists were either clinical psychology doctoral trainees or licensed clinical psychologists. At posttreatment, 64.00% of the youth in the original sample no longer met diagnostic criteria for their principal anxiety disorder based on the Anxiety Disorders Interview Schedule for DSM-IV.

Practice Settings

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. Thirty-eight youth participants (M age = 10.83, SD = 2.12; 36.80% Caucasian; 42.10% male) received treatment at community health clinics across Los Angeles county. All therapists were clinic employees (n = 29) who volunteered to participate in the study and were randomly assigned to treatment groups. Therapists assigned to YAS-ICBT (n = 13; 15.40% male) were

Table 1
Youth Descriptive Data and Comparisons Across Groups

Variable	M (SD) or %			F or χ^2	p
	ICBT (n = 51)	YAS-ICBT (n = 17)	YAS-UC (n = 21)		
Age	10.36 (1.90)	11.32 (2.32)	10.44 (1.91)	1.56	.217
Sex					
Male	60.80 ^e	29.40	52.40 ^f	5.04	.081
Race/Ethnicity				29.91	<.001
Caucasian	86.20 ^a	41.20	33.30		
African-American	9.80	-	9.50		
Latino	2.00	17.60 ^b	42.90 ^c		
Mixed/Other	2.00	5.90	9.50		
Not Reported	-	35.30 ^d	4.80		
CBCL					
Total	63.18 (8.44)	64.19 (7.34)	65.06 (6.46)	0.39	.678
Internalizing	67.40 (8.37)	66.38 (8.33)	66.82 (8.33)	0.10	.904
Externalizing	52.96 (10.08)	60.81 (7.49) ^b	59.41 (9.67)	5.61	.005
Anxious-Depressed	63.10 (9.31)	68.00 (8.69) ^b	68.29 (8.37) ^c	3.51	.034
Primary Diagnoses				22.73	.001
GAD	37.30 ^a	5.90	14.30		
SAD	29.40	35.30	38.10		
SOP	33.30	23.50	28.60		
SP	-	35.30 ^b	19.00 ^c		
Family Income				15.66	<.001
Up to 60k per year	35.30	70.60 ^b	76.20 ^c		
Number of Sessions	15.92 (1.43)	16.82 (5.02)	15.71 (9.34)	0.26	.775
Weeks in Treatment	19.52 (3.97)	26.38 (10.41) ^b	26.84 (15.53) ^c	6.45	.002
Number of Coded Sessions	10.43 (2.84)	12.47 (4.61)	10.00 (6.00)	1.94	.145

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. CBCL = Child Behavior Checklist, GAD = generalized anxiety disorder, SAD = separation anxiety disorder, SOP = social phobia, SP = specific phobia. Analysis of variance was conducted with continuous variables whereas chi-square analyses were conducted with categorical variables.

^a = ICBT > YAS-ICBT, YAS-UC.

^b = YAS-ICBT > ICBT.

^c = YAS-UC > ICBT.

^d = YAS-ICBT > ICBT, YAS-UC.

^e = ICBT > YAS-ICBT.

^f = YAS-UC > YAS-ICBT.

53.80% Caucasian, 15.40% Latino, 15.40% Asian/Pacific Islander, and 15.40% mixed/other. Therapists assigned to YAS-UC ($n = 16$; 12.50% male) were 43.70% Caucasian, 37.50% Latino, and 12.50% mixed/other (6.30% did not report). Professional composition of YAS-ICBT therapists was 30.80% social workers, 23.10% master's-level psychologists, 15.30% doctoral-level psychologists, and 30.80% reported "other" degree (e.g., marital and family therapists, master's of counseling). YAS-ICBT therapists endorsed a variety of theoretical orientations including 38.40% psychodynamic, 30.80% cognitive-behavioral, 15.40% family systems, and 15.40% "other." In YAS-UC, therapists were 25.00% social workers, 31.20% master's-level psychologists, 6.30% doctoral-level psychologists, 31.20% were "other," and 6.30% did not report. The YAS-UC therapists reported a variety of theoret-

ical orientations including 31.30% psychodynamic, 6.30% cognitive-behavioral, 18.70% family systems, 18.70% eclectic, and 6.30% "other" (6.30% did not report). At posttreatment, 66.70% and 73.70% of youths in the original study no longer met diagnostic criteria based on the Diagnostic Interview Schedule for Children 4.0 in the YAS-ICBT and YAS-UC groups, respectively.

TREATMENTS

ICBT

Therapists in ICBT and YAS-ICBT delivered *Coping Cat*, an ICBT program for youth diagnosed with anxiety disorders (Kendall & Hedtke, 2006). *Coping Cat* consists of 16 sessions (14 with the youth; 2 with parents). The program emphasizes anxiety management skills training (e.g., cognitive restructuring,

relaxation), exposure, and regular homework assignment. The *Coping Cat* manual also consistently encourages the use of client-centered interventions, such as portraying warmth and active listening. In both studies, the therapists were trained using the same procedure: studying the treatment manual, engaging in a training workshop, and attending weekly supervision with a model expert (i.e., therapists in each study were supervised by an individual trained by the program developer). Scores on the Coping Cat Brief Adherence Scale (see Kendall, 1994) indicated that adherence was high in ICBT and YAS-ICBT (> 90% of expected content). A treatment differentiation check that used the Therapy Process Observational Coding System for Child Psychotherapy–Strategies scale (McLeod & Weisz, 2010) determined that the YAS-ICBT and YAS-UC groups did not overlap significantly in ICBT interventions unique to *Coping Cat* whereas YAS-UC scored higher on psychodynamic and family interventions (see Kendall et al., 2008; Southam-Gerow et al., 2010, for details).

Usual Care (UC)

Therapists who delivered UC agreed to use the therapeutic interventions they regularly provided and believed to be effective in their routine practice. The UC therapists received clinical supervision provided as standard practice in their clinic.

THERAPY PROCESS INSTRUMENT

Therapy Process Observational Coding System for Child Psychotherapy–Revised Strategies Scale (TPOCS-RS; McLeod, Smith, Southam-Gerow, Weisz, & Kendall, 2016)

The 42-item TPOCS-RS is an observer-based instrument that assesses the extent that interventions are delivered within 5 theory-based subscales: Cognitive (4 items), Behavioral (9 items), Psychodynamic (5 items), Client-Centered (4 items), and Family (7 items). The instrument also contains 13 general items (e.g., *Homework, Play Therapy*) that represent interventions considered to play a meaningful role in treatment but are not associated with a specific theory-based subscale (Orlinsky, Rønnestad, & Willutzki, 2004; Weersing, Weisz, & Donenberg, 2002). The scoring strategy includes extensiveness ratings to assess 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 uses each intervention during the entire session using a 7-point Likert-type extensiveness scale (Hogue, Liddle, & Rowe, 1996) 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 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 et al., 1996). A coder considers both thoroughness and frequency while making a rating; extensiveness ratings thus provide dosage information about each intervention. Previous studies have demonstrated that the TPOCS-RS, or variants thereof (i.e., PRAC-TPOCS, Garland et al., 2010; TPOCS-S, McLeod & Weisz, 2010), have demonstrated item interrater reliability ranging from .71 to .86 ($M ICC = .81$), the item and subscale scores provide evidence of construct validity across research and practice settings (McLeod & Weisz, 2010; McLeod et al., 2015; Southam-Gerow et al., 2016; Wood et al., 2006), subscale scores differentiate between treatment types (McLeod et al., 2015; Southam-Gerow et al., 2010; Weisz et al., 2009; Wood et al., 2006), and subscale scores demonstrated predictive validity (Garland et al., 2014). The current study used the TPOCS-RS Psychodynamic, Family, and Client-centered subscales. We also used the TPOCS-RS Coping Cat subscale, which is comprised of items from the Cognitive and Behavioral subscales that match the content in the *Coping Cat* program (i.e., *Relaxation, Cognitive Education, Cognitive Distortion, Coping Skills, Operant, Respondent*; see Southam-Gerow et al., 2010). We calculated interrater reliability for each subscale ($N = 4$) included in the analyses using ICC (Shrout & Fleiss, 1979). The reliability coefficients represent the model ICC(2, 2) based on a two-way random effects model; subscale interrater reliability ICCs ranged from .72 to .94 ($M = .86$, $SD = .08$): Coping Cat ICC = .89, Psychodynamic ICC = .89, Family ICC = .94, and Client-centered ICC = .72. 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.

TPOCS-RS CODING AND SESSION SAMPLING PROCEDURES

Two doctoral student coders, naïve to study group, were trained over a 3-month period. Coder training involved three steps. First, coders received didactic instruction and discussion of the scoring manual, reviewed sessions with the trainers (second and third authors), 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 (second and third authors). To be certified for independent coding, each coder

had to demonstrate “good” reliability on each item ($ICC(2, 2) > .59$; Cicchetti, 1994). After coders were certified for coding, recordings were randomly assigned. During coding, regular meetings were held and interrater reliability was assessed monthly to prevent coder drift. Each session was double coded and the mean score was used in analyses to reduce measurement error.

All sessions for each case were coded except for the first and last session, as these sessions may contain intake or termination content. Existing recordings were not rated if (a) audible content was shorter than 15 minutes, (b) less than 75% of the dialogue was in English, or (c) missing or damaged. Of the 1,428 sessions held, 983 (68.8%) were recorded and 445 (31.2%) were not recorded. Of the recorded sessions, 954 (97%) were rated by both coders (66% ICBT [$n = 532$], 75% YAS-ICBT [$n = 212$], 67% YAS-UC [$n = 210$]) and 29 (3%) were not coded ($n = 25$ sessions were less than 15 minutes; $n = 4$ sessions the dialogue was less than 75% English). There were no significant differences across groups in terms of the reason why sessions were not coded, the percent of sessions coded, or the percent of sessions coded from the first and second half of treatment.

ASSESSMENT INSTRUMENTS COLLECTED IN THE ORIGINAL STUDIES

The Kendall et al. (2008) study used the *Anxiety Disorders Interview Schedule for DSM-IV: Child and Parent Versions* (ADIS-C/P; Silverman & Albano, 1996) to assess youth DSM-IV disorders based on separate interviews with the caregivers and the youth. The YAS used the *Diagnostic Interview Schedule for Children Version 4.0* (DISC 4.0; Shaffer, Fisher, Dulcan, & Davies, 1996) to assess youth DSM-IV disorders based on separate interviews with the caregivers and the youth. Both studies collected the *Child Behavior Checklist* (CBCL; Achenbach, 1991) to assess symptoms across a broad range of clinical significance (e.g., shyness to suicide attempts). CBCL raw scores are converted to *T*-scores for three broad-band scales (e.g., Internalizing) and eight narrow-band subscales (e.g., Somatic Complaints). In the current investigation, three CBCL scales were used: Internalizing (broad-band), Externalizing (broad-band), and Anxious-Depressed (narrow-band). See Kendall et al. (2008) and Southam-Gerow et al. (2010) for more information.

DATA ANALYSIS

We first evaluated missing data patterns. As noted above, the number of sessions coded did not vary across groups so these analyses focused on patterns of missingness in youth-level control variables and

TPOCS-RS subscale data. Rates of missing data were 9.00% (income data) or less across variables. These data were missing completely at random (Little’s MCAR test $\chi^2 = 159.776$, $df = 144$, $p = .174$). Due to the missing data, subsequent analyses involving the youth-level variables were conducted using the multiple imputation function in HLM 7.01, using 10 datasets imputed using IBM SPSS Statistics version 22.0.

We calculated TPOCS-RS subscale scores. Without consensus on how to score integrity instruments (see McLeod et al., 2015; Southam-Gerow et al., 2016), we considered three approaches: (a) Average of all items on each subscale for each session (Coping Cat $M = 2.63$, $SD = 1.22$; Psychodynamic $M = 1.19$, $SD = 0.35$; Family $M = 1.57$, $SD = 0.85$; Client-Centered $M = 2.76$, $SD = 0.76$); (b) Average of all scored items (items scored above a 1) for each session (Coping Cat $M = 3.08$, $SD = 1.22$; Psychodynamic $M = 1.40$, $SD = 0.60$; Family $M = 1.96$, $SD = 1.14$; Client-Centered $M = 3.11$, $SD = 0.79$); and (c) Highest item scored for each session (Coping Cat $M = 4.56$, $SD = 2.10$; Psychodynamic $M = 1.49$, $SD = 0.81$; Family $M = 2.52$, $SD = 1.91$; Client-Centered $M = 4.28$, $SD = 1.41$). Though scores on the three scoring approaches differed in magnitude, the three scores for each subscale were highly correlated (all $r_s > .78$). We chose to use the highest score on each subscale so that the scores were easily translated to the 1–7 rating approach used for the measure (cf. Southam-Gerow et al., 2016).

Analyses of group differences were conducted using multilevel modeling (Raudenbush & Bryk, 2002) with HLM 7.01 (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011) to account for the nesting of sessions within youth and youth nested within therapists. Given findings suggesting intervention delivery may change over time (e.g., Boswell et al., 2013), unconditional 3-level growth models were fit to each TPOCS-RS subscale (Coping Cat, Psychodynamic, Client-Centered, Family) to examine whether change was linear, quadratic, or loglinear. Time was centered at the first session to assess group differences at the beginning of treatment.

Group differences in TPOCS-RS subscale scores were evaluated by dummy coding group membership and entering it as a youth-level (level 2) predictor of intercept (subscale score at first session) and slope (e.g., type of change in subscale scores over time). The models for each TPOCS-RS subscale score were run twice, once with YAS-UC as the reference group and once with ICBT as the reference group, enabling all pairwise comparisons. Below is an example of a 3-level model that was fit to the Coping Cat subscale

scores, in which change over time is represented by a quadratic term and YAS-UC is the reference group:

$$\text{L1: } \text{COPINGCA}_{ijk} = \pi_{0jk} + \pi_{1jk} * (\text{TIMEWEEK}_{ijk}) \\ + \pi_{2jk} * (\text{WEEKSSQ}_{ijk}) + e_{ijk}$$

$$\text{L2: } \pi_{0jk} = \beta_{00k} + \beta_{01k} * (\text{ICBT}_{jk}) \\ + \beta_{02k} * (\text{YAS-ICBT}_{jk}) + r_{0jk} \\ \pi_{1jk} = \beta_{10k} + \beta_{11k} * (\text{ICBT}_{jk}) \\ + \beta_{12k} * (\text{YAS-ICBT}_{jk}) + r_{1jk} \\ \pi_{2jk} = \beta_{20k} + \beta_{21k} * (\text{ICBT}_{jk}) \\ + \beta_{22k} * (\text{YAS-ICBT}_{jk}) + r_{2jk}$$

$$\text{L3: } \beta_{00k} = \gamma_{000} + u_{00k} \\ \beta_{01k} = \gamma_{010} \\ \beta_{02k} = \gamma_{020} \\ \beta_{03k} = \gamma_{030} \\ \beta_{10k} = \gamma_{100} \\ \beta_{11k} = \gamma_{110} \\ \beta_{12k} = \gamma_{120} \\ \beta_{13k} = \gamma_{130}$$

Cohen's *d* effect sizes and 95% CIs were calculated. For each subscale, group differences were computed at the intercept level (i.e., group mean differences) by dividing the group difference parameter estimate by the raw data SDs. For the Coping Cat and Psychodynamic subscale scores, group differences in slope were also assessed for both linear and quadratic models. Effect sizes were computed following Feingold's (2009) recommendation to multiply the parameter by the length of treatment and divide it by the raw data SD. Given that treatment length varied across groups (see Table 1), we used the average length of treatment of 23 weeks. In models with quadratic trends, Feingold's formula was modified to incorporate the linear and quadratic parameters. All of these effect size calculations represent Cohen's (1988) *d*, in which .2 is considered small, .5 medium, and .8 large.

Because differences in youth demographic and clinical characteristics may explain differences in treatment delivery across settings, we examined whether the group differences on the TPOCS-RS subscales remained when controlling for the youth demographic and clinical characteristics that differed between the groups (see Table 1). As noted above, the analyses with youth variables were conducted using the multiple imputation function in HLM 7.01.

Results

We first conducted sample bias analyses to examine whether our subsamples represented the full samples from the parent studies. One significant difference was found. Compared to the parent study, there was a lower percentage of African-American youth (0.00% vs. 16.70%) and a higher percentage of Caucasian youth (41.20% vs. 29.20%) in our YAS-ICBT group, $\chi^2(3, n = 24) = 11.53, p = .009$. As reported in Table 1, we also examined differences between groups (ICBT, YAS-ICBT, YAS-UC) on youth demographic, baseline clinical, and treatment characteristics as well as therapist characteristics (sex, race/ethnicity, theoretical orientation). To be conservative, we set the significance criterion at $p < .20$. The groups differed on sex, race/ethnicity (categorized as minority vs. Caucasian), level of externalizing symptomatology, level of anxiety and depressive symptomatology, primary anxiety disorders (i.e., specific phobia and generalized anxiety disorder), family income level, and weeks in treatment.

Mean scores for the TPOCS-RS subscales for each group are presented in Table 2. Paired *t*-tests were used to evaluate subscale scores within each group (see Figure 1) using a Bonferroni adjusted alpha of .008. For ICBT, Coping Cat scores were significantly higher than the Client-Centered, ($t = 13.23, p < .001, d = 0.66, 95\% \text{ CI } [0.53, 0.78]$), Family ($t = 30.48; p < .001, d = 2.27, 95\% \text{ CI } [2.12, 2.42]$), and Psychodynamic ($t = 60.32; p < .001, d = 3.84, 95\% \text{ CI } [3.63, 4.04]$) scores; Client-Centered scores were significantly higher than Family ($t = 24.92; p < .001, d = 1.71, 95\% \text{ CI } [1.57, 1.85]$) and Psychodynamic ($t = 49.59; p < .001, d = 3.17, 95\%$

Table 2
Means, Standard Deviations, and Group Differences Across TPOCS-RS Subscales

TPOCS-RS Subscale	ICBT		YAS-ICBT		YAS-UC	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Coping Cat	5.66 ^a	1.54	4.34 ^b	1.78	2.01	1.12
Psychodynamic	1.23	0.54	1.59 ^c	0.83	2.05 ^d	1.03
Family	2.00	1.68	2.42	1.73	3.93 ^d	1.95
Client-centered	4.68 ^a	1.44	3.82	1.30	3.73	1.08

Note. TPOCS-RS = The Therapy Process Observational Coding System for Child Psychotherapy – Revised Strategies scale; ICBT = individual cognitive-behavioral treatment delivered in Kendall et al. study; YAS-ICBT = ICBT delivered in YAS; YAS-UC = usual care delivered in YAS. The TPOCS-RS scores were produced by calculating the mean scores for all sessions that the youth attended from each case.

^a = ICBT > YAS-ICBT, YAS-UC

^b = YAS-ICBT > YAS-UC

^c = YAS-ICBT > ICBT

^d = YAS-UC > ICBT, YAS-ICBT

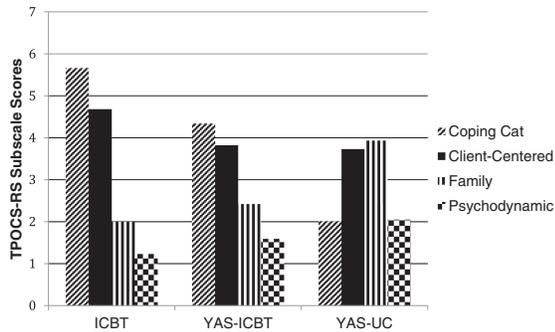


FIGURE 1 Average TPOCS-RS scores by group and subscale. TPOCS-RS = The Therapy Process Observational Coding System for Child Psychotherapy – Revised Strategies scale; ICBT = individual cognitive-behavioral treatment delivered in Kendall et al. study; YAS-ICBT = ICBT delivered in YAS; YAS-UC = usual care delivered in YAS.

CI [2.99, 3.35]) scores; Family scores were significantly higher than Psychodynamic scores ($t = 10.21$; $p < .001$, $d = 0.62$, 95% CI [0.49, 0.74]). For YAS-ICBT, Coping Cat scores were significantly higher than Client-Centered ($t = 4.14$, $p < .001$, $d = 0.33$, 95% CI [0.14, 0.52]), Family ($t = 9.64$; $p < .001$, $d = 1.09$, 95% CI [0.89, 1.30]), and Psychodynamic ($t = 18.08$; $p < .001$, $d = 1.98$, 95% CI [1.74, 2.21]) scores; Client-Centered scores were significantly higher than Family ($t = 8.81$; $p < .001$, $d = 0.91$, 95% CI [0.71, 1.11]) and Psychodynamic ($t = 20.34$; $p < .001$, $d = 2.04$, 95% CI [1.81, 2.28]) scores; and Family scores were significantly higher than Psychodynamic ($t = 6.95$; $p < .001$, $d = 0.61$, 95% CI [0.42, 0.81]) scores. In YAS-UC, Family scores were significantly higher than Coping Cat ($t = 12.61$, $p < .001$, $d = 1.21$, 95% CI [1.00, 1.41]) and Psychodynamic ($t = 6.91$, $p < .001$, $d = 1.21$, 95% CI [1.01, 1.42]) scores; Client-Centered scores were significantly higher than Coping Cat ($t = 17.52$ $p < .001$, $d = 1.56$, 95% CI [1.34, 1.78]) and Psychodynamic ($t = 16.72$ $p < .001$, $d = 1.59$, 95% CI [1.37, 1.88]) scores; there were no significant differences between Family and Client-Centered scores ($t = -1.36$, $p = 0.174$, $d = 0.13$, 95% CI [0.00, 0.32]) or Coping Cat and Psychodynamic scores ($t = 0.32$, $p = 0.729$, $d = 0.04$, 95% CI [0.00, 0.15]).¹

HLM ANALYSES

We tested for group differences by building models for each TPOCS-RS subscale: Coping Cat, Psychodynamic, Family, and Client-centered. For the

Coping Cat and Psychodynamic subscales, adding a linear term improved model fit relative to the intercept-only model (Coping Cat Linear Deviance Difference $\chi^2 = 7.8$, $df = 3$, $p < .05$; Psychodynamic Linear Deviance Difference $\chi^2 = 57.01$, $df = 3$, $p < .05$); adding a quadratic term further improved model fit (Deviance Difference Coping Cat Quadratic $\chi^2 = 16.62$, $df = 4$, $p < .05$; Deviance Difference Psychodynamic Quadratic $\chi^2 = 13.90$, $df = 4$, $p < .05$). Entering a loglinear term to Coping Cat and Psychodynamic models did not improve model fit relative to the intercept-only model, so the data for these two subscales appeared to fit quadratic models. For the Family and Client-centered subscales, adding a linear term did not improve model fit (Deviance Difference Family $\chi^2 = 4.69$, $df = 3$, $p < .05$; Deviance Difference Client-centered $\chi^2 = 1.78$, $df = 3$, $p < .05$); nor did adding quadratic or loglinear terms. Thus, an intercept-only model appeared to be the best fit for these subscales. Residual files from these models were next used to identify outliers (i.e., data more than 3 SDs from the mean; Van Selst & Jolicoeur, 1994) on intercept and slope. Across all subscales, 4 cases were outliers on intercept (1 on the Family subscale; 3 on the Psychodynamic subscale) and 2 were slope outliers on the Psychodynamic subscale. Two of the four cases (both in UC) had both intercept and slope outliers on the Psychodynamic subscale raising concerns about their potential to impact study findings, so analyses were run with and without these cases (see below).²

Coping Cat Subscale

Model results are presented in Table 3. Across all groups, the Coping Cat subscale score was 3.60 (i.e., the intercept; $p < .001$) at the first session. Coping Cat subscale scores changed over treatment (Coping Cat $y_{100} = 0.104$; $p < .001$ and $y_{200} = -0.004$; $p < .001$). In the base model (i.e., without the group difference predictors), 40.43% of the variability in intercept was at the therapist level, 9.88% at the youth level, and 55.25% at the session level. Over 99.00% of the variability in quadratic slope was at the session level, so all slope effects were fixed at level 3 in subsequent models. As hypothesized, at the first session, Coping Cat subscale scores in the 2 ICBT groups were higher than YAS-UC (ICBT vs. YAS-UC $\gamma = 2.62$, $p < .001$, $d = 3.16$, 95% CI [2.77, 3.54]; YAS-ICBT vs. YAS-UC $\gamma = 2.54$, $p < .001$, $d = 3.06$, 95% CI [2.62, 3.49]). Contrary to hypotheses, the ICBT groups did not significantly differ on Coping Cat subscale scores at the first session (ICBT vs. YAS-ICBT $\gamma =$

¹The same pattern held across the four TPOCS-RS subscales when we used the two other subscale scoring methods (see Supplemental Table 1 in online materials).

²We reran the HLM models with the two alternate scoring approaches and found the same pattern of findings (see Supplemental Table 2 in online materials).

Table 3
Multilevel Models Comparing TPOCS-RS Subscale Scores Across Groups

	Coefficient	S.E.	ES 95% [CI]	Deviance	<i>n</i> Parameters in Model
Coping Cat Subscale					
Base Quadratic Model				3479.81	11
Intercept (first session value), Y_{000}	3.60 ***	0.28	N/A		
Slope (change over time in weeks), Y_{100}	0.104 ***	0.02	N/A		
Slope (change over time in weeks squared), Y_{200}	-0.004 ***	0.001	N/A		
Models including group contrasts ¹				3368.54	17
Intercept (first session value), Y_{000}	4.65 ***	0.21	N/A		
Slope (change over time in weeks), Y_{100}	0.20 ***	0.03	N/A		
Slope (change over time in weeks squared), Y_{200}	-0.01 ***	0.002	N/A		
Group Differences in Intercept					
ICBT (1) vs. YAS-UC (0)	2.62 ***	0.32	3.16 [2.77, 3.54]		
YAS-ICBT (1) vs. YAS-UC (0)	2.54 ***	0.36	3.06 [2.62, 3.49]		
ICBT (1) vs. YAS-ICBT (0)	0.08	0.34	0.10 [0.00, 0.51]		
Group Differences in Linear Slope, Y_{110}, Y_{120}					
ICBT (1) vs. YAS-UC (0)	0.20 ***	0.04	5.54 [4.43, 6.65]		
YAS-ICBT (1) vs. YAS-UC (0)	0.02	0.04	0.55 [0.00, 1.66]		
ICBT (1) vs. YAS-ICBT (0)	0.17 **	0.05	4.82 [5.54, 6.09]		
Group Differences in Quadratic Slope, Y_{210}, Y_{220}					
ICBT (1) vs. YAS-UC (0)	-0.01 ***	0.002			
YAS-ICBT (1) vs. YAS-UC (0)	-0.002	0.001			
ICBT (1) vs. YAS-ICBT (0)	-0.01 *	0.002			
Psychodynamic Subscale					
Base Quadratic Model				1898.14	11
Intercept (first session value), Y_{00}	1.46 ***	0.10	N/A		
Slope (change over time in weeks), Y_{100}	0.006	0.01	N/A		
Slope (change over time in weeks squared), Y_{200}	0.0003	0.0006	N/A		
Models including group contrasts ¹				1864.54	17
Intercept (first session value), Y_{00}	2.09 ***	0.15	N/A		
Slope (change over time in weeks), Y_{100}	-0.03	0.03	N/A		
Slope (change over time in weeks squared), Y_{200}	0.002 *	0.001	N/A		
Group Differences in Intercept					
ICBT (1) vs. YAS-UC (0)	-0.94 ***	0.18	1.54 [1.25, 1.84]		
YAS-ICBT (1) vs. YAS-UC (0)	-0.87 ***	0.22	1.41 [1.07, 1.80]		
ICBT (1) vs. YAS-ICBT (0)	-0.08	0.19	0.13 [0.00, 0.45]		
Group Differences in Linear Slope, Y_{110}, Y_{120}					
ICBT (1) vs. YAS-UC (0)	0.03	0.03	1.13 [0.00, 2.27]		
YAS-ICBT (1) vs. YAS-UC (0)	0.09 **	0.03	3.39 [2.27, 4.55]		
ICBT (1) vs. YAS-ICBT (0)	-0.06	0.03	2.26 [1.14, 3.42]		
Group Differences in Quadratic Slope, Y_{210}, Y_{220}					
ICBT (1) vs. YAS-UC (0)	-0.001	0.001			
YAS-ICBT (1) vs. YAS-UC (0)	-0.003 *	0.001			
ICBT (1) vs. YAS-ICBT (0)	-0.002	0.001			
Family Subscale					
Unconditional Means Model				3809.29	4
Intercept (first session value), Y_{00}	2.79 ***	0.17	N/A		
Models including group contrasts ¹				3764.19	6
Intercept (first session value), Y_{00}	2.01 ***	0.14	N/A		
Group Differences in Intercept					
ICBT (1) vs. YAS-UC (0)	-1.96 ***	0.22	1.03 [0.91, 1.14]		
YAS-ICBT (1) vs. YAS-UC (0)	-1.58 ***	0.24	0.83 [0.70, 0.95]		
ICBT (1) vs. YAS-ICBT (0)	-0.38	0.22	0.20 [0.08, 0.31]		
Client-centered Subscale					
Unconditional Means Model				3038.57	4
Intercept (first session value), Y_{00}	4.16 ***	0.13	N/A		

(continued on next page)

Table 3 (continued)

	Coefficient	S.E.	ES 95% [CI]	Deviance	<i>n</i> Parameters in Model
Models including group contrasts ¹				3028.64	6
Intercept (first session value), γ_{00}	4.65 ***	0.19	N/A		
Group Differences in Intercept					
ICBT (1) vs. YAS-UC (0)	0.88 **	0.28	0.62 [0.43, 0.82]		
YAS-ICBT (1) vs. YAS-UC (0)	0.18	0.30	0.13 [0.00, 0.34]		
ICBT (1) vs. YAS-ICBT (0)	0.70 *	0.29	0.50 [0.29, 0.70]		

Note. For parsimony, this table includes only parameters of central conceptual interest (e.g., intercepts are not reported for the models examining group differences, parameters are not reported for the control variables).

TPOCS-RS = The Therapy Process Observational Coding System for Child Psychotherapy – Revised Strategies scale; S.E. = Standard Error; ES = effect size (Cohen's *d*); ICBT = individual cognitive-behavioral treatment delivered in Kendall et al. study; YAS-ICBT = ICBT delivered in YAS; YAS-UC = usual care delivered in YAS.

* $p < .05$;

** $p < .01$;

*** $p < .001$

0.08, $p = 0.821$, $d = 0.10$, 95% CI [0.00, 0.51]). Over time, the ICBT group differed from the YAS-ICBT and YAS-UC groups in both linear and quadratic changes (see Figure 2). ICBT had significant quadratic and linear trends; the trajectory increased until the end of treatment when the trajectory decreased slightly. In contrast, the two YAS groups did not differ significantly from each other over time as each group evidenced a flat trajectory. Because the groups evidenced different trajectories, we conducted post-hoc analyses to determine if group differences existed as treatment progressed by re-centering the subscale at a middle session and late session. Consistent with our hypotheses, Coping Cat subscale scores were signif-

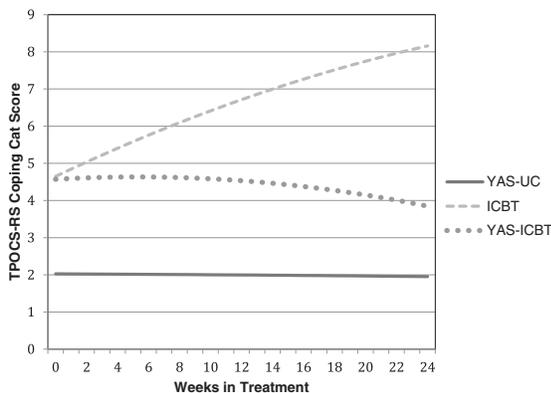


FIGURE 2 TPOCS-RS Coping Cat scores over time by group. TPOCS-RS = The Therapy Process Observational Coding System for Child Psychotherapy – Revised Strategies scale; ICBT = individual cognitive-behavioral treatment delivered in Kendall et al. study; YAS-ICBT = ICBT delivered in YAS; YAS-UC = usual care delivered in YAS.

icantly higher in ICBT than YAS-ICBT at a middle session (i.e., week 8; $\gamma = 1.06$, $p < .001$, $d = 0.52$, 95% CI [0.39, 0.64]) and at a late session (i.e., week 16; $\gamma = 1.81$, $p < .001$, $d = 0.87$, 95% CI [0.73, 1.01]). At each time interval, the Coping Cat subscale scores remained higher in both ICBT groups than YAS-UC.

Psychodynamic Subscale

Across all groups, the Psychodynamic subscale score averaged 1.46 (i.e., the intercept; $p < .001$) at the first session. Psychodynamic scores did not change significantly over the course of treatment (Psychodynamic Linear Slope $\gamma_{100} = .0006$; $p = .654$ and Quadratic Slope $\gamma_{200} = .0003$; $p = .523$). In the base quadratic model, 24.66% of the variability in intercept was at the therapist level, 31.50% was at the youth level, and 43.84% was at the session level. Over 99.00% of the variability in quadratic slope was at the session level so in subsequent models all slope effects were fixed at level 3. As hypothesized, at the first session, Psychodynamic subscale scores were higher in UC than both ICBT groups (YAS-UC vs. ICBT $\gamma = 0.94$, $p < .001$, $d = 1.54$, 95% CI [1.25, 1.84]; YAS-UC vs. YAS-ICBT $\gamma = 0.87$, $p < .001$, $d = 1.41$, 95% CI [1.07, 1.80]). However, contrary to our hypotheses, scores were similar in both ICBT groups (ICBT vs. YAS-ICBT $\gamma = -0.08$, $p = .684$, $d = 0.13$, 95% CI [0.00, 0.45]). Over time, therapists in the two YAS groups differed in both their linear and quadratic changes.³ YAS-UC had a significant quadratic trend, but not

³ Group comparisons on the Psychodynamic subscale were rerun with the two cases that had the intercept and slope outliers. All findings remained the same except that the two YAS groups no longer differed in quadratic changes (YAS-ICBT vs. YAS-UC $\gamma_{200} = -.002$, $p = .077$, $d = 0.63$).

a linear one; the trajectory dipped slightly mid-treatment, but returned to similar levels by the end of treatment (see Figure 3). In contrast, YAS-ICBT scores rose steadily until mid-treatment and then tapered off. ICBT trajectories did not differ significantly from the other groups; the trajectory remained relatively flat, rising slightly at the end of treatment. Post-hoc analyses, re-centering the Psychodynamic subscale at a middle session and late session, indicated that the Psychodynamic subscale scores were higher in YAS-ICBT than ICBT at a middle (i.e., week 8; $\gamma = 0.35, p = .005, d = 0.85, 95\% \text{ CI } [0.56, 1.15]$) and at a late session (i.e., week 16; $\gamma = 0.37, p = .019, d = 0.21, 95\% \text{ CI } [0.12, 0.29]$). Psychodynamic subscale scores were higher in YAS-UC than YAS-ICBT at a middle session ($\gamma = 0.47, p = .001, d = 1.15, 95\% \text{ CI } [0.80, 1.49]$), but not at a late session ($\gamma = 0.34, p = .061, d = 0.18, 95\% \text{ CI } [0.09, 0.47]$).

Client-Centered Subscale

Client-Centered subscale scores averaged 4.16 over treatment (i.e., the intercept; $p < .001$). Scores were significantly higher in ICBT than in YAS-ICBT ($\gamma = .70, p = .020, d = 0.50, 95\% \text{ CI } [0.29, 0.70]$) and YAS-UC ($\gamma = 0.88, p = .003, d = 0.62, 95\% \text{ CI } [0.43, 0.82]$). Scores were not significantly different in the YAS groups (YAS-ICBT vs. YAS-UC $\gamma = .18, p = .554, d = 0.13, 95\% \text{ CI } [0.00, 0.34]$).

Family Subscale

Family subscale scores averaged 2.79 (i.e., the intercept; $p < .001$). Scores were significantly higher in UC than both ICBT groups (YAS-UC vs. ICBT $\gamma = 1.96, p < .001, d = 1.03, 95\% \text{ CI } [0.91, 1.14]$; YAS-UC vs. YAS-ICBT $\gamma = 1.58, p < .001, d = 0.83, 95\% \text{ CI } [0.70, 0.95]$), but the scores were not significantly different in the ICBT groups (ICBT vs. YAS-ICBT $\gamma = -0.38, p = .097, d = 0.20, 95\% \text{ CI } [0.08, 0.31]$).

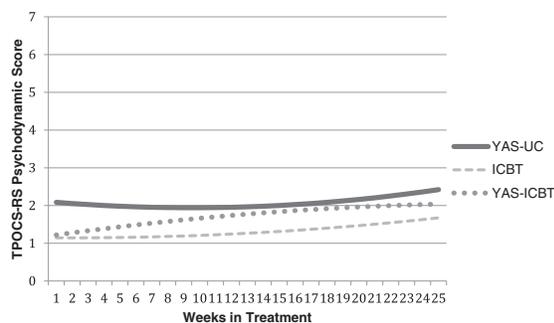


FIGURE 3 TPOCS-RS psychodynamic scores over time by group. TPOCS-RS = The Therapy Process Observational Coding System for Child Psychotherapy – Revised Strategies scale; ICBT = individual cognitive-behavioral treatment delivered in Kendall et al. study; YAS-ICBT = ICBT delivered in YAS; YAS-UC = usual care delivered in YAS.

RULING OUT ALTERNATIVE INTERPRETATIONS

We examined whether findings held when controlling for characteristics that differed across groups (see Table 1). Control variables were entered simultaneously into each subscale model, using multiple imputation to account for missing data across groups: sex, race/ethnicity (categorized as minority vs. Caucasian), level of externalizing symptomatology, level of anxiety and depressive symptomatology, primary anxiety disorders (i.e., specific phobia and generalized anxiety disorder), family income level, and weeks in treatment. No findings were altered for any TPOCS-RS subscale.

Discussion

Does the delivery of the same treatment program, ICBT for youth anxiety, differ across research and practice settings? Therapists trained and supervised in the same ICBT program across both settings delivered similar patterns of intervention dosages (i.e., delivered strongest dosage of cognitive-behavioral interventions, followed by client-centered, family, and psychodynamic interventions), whereas therapists not trained in the ICBT program and providing UC delivered a distinct pattern of interventions (i.e., delivered strongest dosages of client-centered and family interventions, lowest dose of cognitive-behavioral interventions). In both settings, therapists trained in ICBT delivered comparable levels of cognitive-behavioral interventions at the start of treatment; however, therapists trained in ICBT in the research setting delivered significantly more cognitive-behavioral interventions at mid- and late-treatment. Over the course of treatment, therapists trained in ICBT in the research setting delivered the strongest dose of client-centered interventions, whereas no difference was observed between the two groups in the practice setting. Regarding the delivery of interventions not found in the protocol (i.e., treatment differentiation), the therapists trained in ICBT in the practice settings delivered more psychodynamic interventions at mid- and late treatment, but no differences were found between the ICBT groups for family-focused interventions. Altogether, our findings indicate that there are more similarities than differences in delivery when the same EBT is delivered across research and practice settings. Moreover, the delivery of the same EBT across settings was largely distinguished from UC in ways predicted by the content of the protocol.

One possible interpretation of these findings is that the delivery of the same EBT does not differ much across settings. When the subscale scores averaged across treatment are considered, this interpretation seems plausible as both ICBT groups displayed the same pattern of scores across the four

TPOCS-RS subscales. Specifically, both ICBT groups delivered interventions considered essential to the *Coping Cat* program (cognitive-behavioral, client-centered) significantly more than proscribed interventions (family, psychodynamic). Also, the scores of the UC group suggested more treatment differentiation—focused on family and client-centered interventions—than either of the two ICBT groups. Thus, when the overall pattern of scores on each of the five theory-driven domains is considered, the two ICBT groups delivered treatment that was (a) consistent with the *Coping Cat* protocol and (b) distinct from UC in predictable ways.

Both studies used a variety of recommended quality-control procedures including a training workshop led by a *Coping Cat* program expert and therapist participation in regular supervision with an expert in the *Coping Cat* program (Beidas & Kendall, 2010; Herschell et al., 2010; Sholomskas et al., 2005). Our data suggest that these quality-control procedures may have worked as intended, increasing the use of *Coping Cat* consistent interventions (i.e., ICBT, client-centered) and limiting the use of interventions not found in the program (i.e., family and psychodynamic), thus resulting in greater treatment differentiation. Our findings are consistent with previous research that has found that quality-control procedures have positively influenced treatment adherence and therapist competence (Sholomskas et al., 2005; Simons et al., 2010).

Though similarities existed between the ICBT groups, there were also some important differences. Two of the four subscale scores (Coping Cat, Psychodynamic) evidenced group differences *in change* over time. At the beginning of treatment, both ICBT groups delivered similar levels of cognitive-behavioral and psychodynamic interventions, but the two groups diverged over the course of treatment. Therapists in the research setting delivered an increasing dose of CBT interventions over time, whereas the dose of CBT interventions for therapists in the practice setting waned over time. In contrast, the therapists trained in ICBT in the research setting delivered a consistent, low dose of psychodynamic interventions across the course of treatment whereas the therapists in the practice setting increased delivery of psychodynamic interventions over time, though the dose remained relatively low. Together, these findings suggest that therapists trained in ICBT in the practice settings displayed lower differentiation as treatment progressed, increasing the dose of non-CBT interventions even as therapists trained in ICBT in the research setting grew stronger in their use of CBT over time.

It is worth considering how and why the amount of cognitive-behavioral interventions delivered by YAS-ICBT therapists waned over time. Therapist-level factors represent one possible reason. Therapists in the research and practice settings differed in potentially significant ways. For example, all therapists in the research setting had a psychology background, whereas only 40% in the practice settings had a psychology background. Moreover, since the research setting specialized in the treatment of anxiety disorders, the therapists in this setting may have had more specialized training. If the therapists in the practice settings did have less experience in delivering cognitive-behavioral interventions, the therapists may have been more reluctant to use certain interventions, such as exposures, that can be difficult to deliver (Levita, Duhne, Girling, & Waller, 2016). It is also possible that as treatment progressed the YAS-ICBT therapists tended to default to more familiar interventions (Becker, Zayfert, & Anderson, 2004; Meyer, Farrell, Kemp, Blakey, & Deacon, 2014). Finally, it is plausible that attitudes towards EBTs, such as ICBT for youth anxiety, may have influenced treatment delivery (Borntrager, Chorpita, Higa-McMillan, & Weisz, 2009; Waller & Turner, 2016), and that such attitudes were more uniformly positive in the research setting. Altogether, future research could evaluate whether therapist-level factors influence the delivery of EBTs.

Although our findings held when differences in youth characteristics were controlled, it is possible that factors not measured herein explain our findings. We only measured a finite number of youth and family-level characteristics. Previous research with adults has found that client factors, such as motivation levels, have been associated with treatment delivery (Boswell et al., 2013; Imel, Baer, Martino, Ball, & Carroll, 2011). Given differences in the referral streams used by the families across settings, it is possible that in our study treatment expectations or motivation may have influenced treatment delivery. Families in the research setting specifically sought out the treatment provided in this specialty setting and thus may have expected to receive ICBT and may have been more motivated to participate than families in the practice settings. Parent treatment expectations are linked with adherence in multisystemic therapy for youth with externalizing disorders (Ellis, Weiss, Han, & Gallop, 2010), so it is plausible that treatment delivery for youth anxiety could be impacted by differences in treatment expectations and motivation across settings. Thus, the potential impact of these, and other, youth- and family-level factors on treatment delivery represents an important avenue for future research.

It is also possible that the changes over time in treatment delivery may reflect adaptations made to the protocol by therapists in the practice settings. Some have suggested the documented differences in youth, therapist, or setting characteristics between research and practice settings necessitate that changes be made to EBT protocols (Weisz et al., 2005). For example, the youth in the practice settings had higher levels of externalizing problems, which may have necessitated the delivery of more family-focused interventions (Chorpita et al., 2011). It is important to note that all three groups produced similar diagnostic outcomes, suggesting that delivering proscribed interventions was not inappropriate for the practice settings (Southam-Gerow et al., 2010). It thus is plausible the observed difference represents adaptations made to the delivery of the ICBT protocol that made it more suitable for practice setting conditions. Of course, it is possible that efforts to increase the dosage of cognitive-behavioral interventions and minimize nonprescribed interventions might have improved outcomes in YAS-ICBT. Indeed, both possibilities represent interesting and important directions for future research.

Our findings pertaining to UC warrant discussion. Previous efforts to describe UC (e.g., Brookman-Frazee, Haine, Baker-Ericzen, Zoffness, & Garland, 2010; Garland et al., 2010; McLeod & Weisz, 2010) have suggested that UC is characterized by breadth (a mix of interventions from multiple theory-based domains) but not depth (low scores). We found that UC included low levels of cognitive, behavioral, and psychodynamic interventions and moderate levels of client-centered and family interventions. The first three findings are consistent with the breadth-not-depth hypothesis, but the latter two are less so. The lack of consistency may be due to our having coded all available sessions, producing a more accurate estimate of dose than past work that sampled a subset of available sessions (e.g., Brookman-Frazee et al., 2010; Garland, et al., 2010; McLeod & Weisz, 2010). Further, our decision to generate TPOCS-RS subscale scores based on the highest item score, as opposed to the average of items (see McLeod & Weisz, 2010), likely led to higher subscale scores compared to past work. We believe our scoring approach is more consistent with how treatment is typically delivered than the alternative, as therapists usually deliver one or two interventions during a single treatment session. Thus, averaging across items may underestimate the actual dosage of interventions delivered. Because our findings challenge the assumption that UC may not contain depth, it is important to highlight this particular methodological issue so that

future researchers can carefully consider different scoring options.

A few limitations of the study bear mentioning. Scores on the Psychodynamic subscale were low across treatment for all three groups, whereas scores on the other three subscales were notably higher. The findings related to the Psychodynamic subscale should thus be interpreted with caution. Another limitation is that, given the nature of efficacy versus effectiveness studies, the Kendall and YAS samples differed on a number of client and therapist characteristics such as primary anxiety disorder and therapists' professional background. Our findings did not change when we controlled for differences across groups; however, our ability to detect differences between groups was somewhat limited by sample size and the fact that the participant information collected across both studies did not include all possible youth, family, and therapist characteristics that might have contributed to group differences. Due in part to sample size, some rather substantial effect sizes were not accompanied by statistically significant group differences. For example, some non-significant group differences in slope (one between YAS groups for Coping Cat scores and one between ICBT and both YAS groups for Psychodynamic scores) produced small to medium effect sizes. Further, because no sample can be completely representative of treatment delivery in research and practice settings, and no single sample can fully characterize UC, it will be important to replicate these comparisons across a wider range of settings, treatments, and youth problems. Finally, it should be noted that the TPOCS-RS was not designed as a treatment adherence instrument. The items do correspond to intervention procedures found in ICBT for youth anxiety, but it is possible that use of an instrument designed specifically to assess adherence to ICBT for youth anxiety might produce different findings.

Our study also has certain strengths. Accurate measurement of treatment delivery was aided by observational data and the use of an instrument that assesses for a wide range of interventions from multiple theory-based domains. The standard in the field is to sample a single session from a few time points for treatment integrity coding (Dennhag, Gibbons, Barber, Gallop, & Crits-Christoph, 2012; Webb, DeRubeis, & Barber, 2010). In contrast, we coded all the available sessions (approximately 11 per case). This sampling not only allowed us to investigate whether treatment delivery systematically increased or decreased over treatment, but provided more precise estimates of treatment delivery at the beginning, middle, and end

of treatment. Investigating differences in delivery over time proved pivotal, given the change in cognitive-behavioral and psychodynamic dosages throughout treatment. Indeed, our results add to a small body of work suggesting that intervention content may change over the course of treatment (e.g., Boswell et al., 2013; Henggeler et al., 2008). And, in fact, our findings suggest that the shape of change (linear, quadratic) for the same theory-based domains may differ across treatments. Together, these findings suggest that relying on very few time points to characterize treatment delivery—the current standard in the field (see Denny et al., 2012)—may miss key information about trends over time in treatment, and indicate that expanded sampling of time points across episodes of care may be warranted in future research.

Some researchers have argued that EBTs may need adaptation or additional therapist training when implemented in practice settings (e.g., Weisz et al., 2005). They have reasoned that pretreatment youth, family, and setting differences across settings may also lead to EBT delivery differences (e.g., adherence, differentiation). The current study suggests that differences in treatment delivery may exist across settings in the context of ICBT for youth anxiety, thus providing some support for these speculations. However, more research is needed to determine what is driving our findings. This comparative characterization study lays a foundation for further investigation into what occurs when EBTs are transported from their research base to the world of clinical practice. Optimizing the delivery of treatments in practice settings represents an important goal for the field. It is possible that bolstering treatment integrity may represent one way to achieve this goal, since treatment integrity is positively linked with clinical outcomes in practice settings (Hogue et al., 2008). Our findings serve to highlight the importance of measuring adherence and differentiation over treatment and suggest that feedback systems that monitor therapist behavior may help identify instances when treatment delivery begins to depart from the protocol. Such information could also be used to guide quality-assurance efforts designed to ensure that EBTs delivered in practice settings are implemented along lines consistent with the treatment model (Bond et al., 2011; Schoenwald, 2011). Future work can use this information to inform quality-assurance efforts by pinpointing the extent to which the delivery of an EBT by therapists in practice settings does, or does not, match the delivery of the same EBT in research settings (McLeod et al., 2013).

Conflict of Interest Statement

The authors declare that there are no conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.beth.2016.07.004>.

References

- Achenbach, T. M. (1991). *Manual for the child behavior checklist 4–18 and 1991 profile*. Burlington, VT: Department of Psychiatry, University of Vermont.
- Bearman, S. K., Weisz, J. R., Chorpita, B. F., Hoagwood, K., Ward, A., Ugueto, A. M., & Bernstein, A. (2013). More practice, less preach? The role of supervision processes and therapist characteristics in EBP implementation. *Administration and Policy in Mental Health and Mental Health Services Research*, 40, 518–529.
- Becker, C. B., Zayfert, C., & Anderson, E. (2004). A survey of psychologists' attitudes towards and utilization of exposure therapy for PTSD. *Behaviour Research and Therapy*, 42, 277–292.
- Beidas, R. S., & Kendall, P. C. (2010). Training therapists in evidence-based practice: A critical review of studies from a systems-contextual perspective. *Clinical Psychology: Science and Practice*, 17, 1–30.
- Bell, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., ... Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH behavior change consortium. *Health Psychology*, 23, 443–451.
- Bond, G. R., Becker, D. R., & Drake, R. E. (2011). Measurement of fidelity of implementation of evidence-based practices: Case example of the IPS Fidelity Scale. *Clinical Psychology: Science and Practice*, 18, 126–141.
- Borntrager, C. F., Chorpita, B. F., Higa-McMillan, C., & Weisz, J. R. (2009). Provider attitudes toward evidence-based practices: Are the concerns with the evidence or the manuals? *Psychiatric Services*, 60, 677–681.
- Boswell, J. F., Gallagher, M. W., Gorman, J. M., Sauer-Zavala, S. E., Bullis, J., Shear, M. K., ... Barlow, D. H. (2013). Patient characteristics and variability in adherence and competence in cognitive-behavioral therapy for panic disorder. *Journal of Consulting and Clinical Psychology*, 81, 443–454.
- Brookman-Frazee, L., Haine, R. A., Baker-Ericzen, M., Zoffness, R., & Garland, A. (2010). Factors associated with evidence-based practice strategies in usual care psychotherapy. *Administration & Policy in Mental Health & Mental Health Services Research*, 37, 254–269.
- Chorpita, B. F., Dalieden, E. L., Ebesutani, C., Young, J., Becker, K. D., Nakamura, B. J., ... Starace, N. (2011). Evidence-based treatments for children and adolescents: An updated review of indicators of efficacy and effectiveness. *Clinical Psychology: Science and Practice*, 18, 154–172.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.) Mahwah, NJ: Lawrence Erlbaum Associates.
- Denny, I., Gibbons, M. B. C., Barber, J. P., 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.
- Ellis, M. L., Weiss, B., Han, S., & Gallop, R. (2010). The influence of parental factors on therapist adherence in multi-systemic therapy. *Journal of Abnormal Child Psychology*, 38, 857–868.

- Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods, 14*(1), 43–53.
- Garland, A. F., Accurso, E. C., Haine-Schlagel, R., Brookman-Frazee, L., Roesch, S., & Zhang, J. J. (2014). Searching for elements of evidence-based practices in children's usual care and examining their impact. *Journal of Clinical Child & Adolescent Psychology, 43*(2), 201–215.
- Garland, A. F., Brookman-Frazee, L., Hurlburt, M. S., Accurso, E. C., Zoffness, R. J., Haine-Schlagel, R., & Ganger, W. (2010). Mental health care for children with disruptive behavior problems: A view inside therapists' offices. *Psychiatric Services, 61*, 788–795.
- Henggeler, S. W., Schoenwald, S. K., Liao, J. G., Letourneau, E. J., & Edwards, D. E. (2002). Transporting efficacious treatments to field settings: The link between supervisory practices and therapist fidelity in MST programs. *Journal of Clinical Child and Adolescent Psychology, 31*, 155–167.
- Henggeler, S. W., Sheidow, A. J., Cunningham, P. B., Donohue, B. C., & Ford, J. D. (2008). Promoting the implementation of an evidence-based intervention for adolescent marijuana abuse in community settings: Testing the use of intensive quality assurance. *Journal of Clinical Child & Adolescent Psychology, 37*, 682–689.
- Herschell, A. D., Kolko, D. J., Baumann, B. L., & Davis, A. C. (2010). The role of therapist training in the implementation of psychosocial treatments: A review and critique with recommendations. *Clinical Psychology Review, 30*, 448–466.
- Hogue, A., Liddle, H. A., & Rowe, C. (1996). Treatment adherence process research in family therapy: A rationale and some practical guidelines. *Psychotherapy: Theory, Research, Practice, Training, 33*, 332–345.
- Hogue, A., Henderson, C. E., Dauber, S., Barajas, P. C., Fried, A., & Liddle, H. A. (2008). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology, 76*(4), 544–555.
- Imel, Z. E., Baer, J. S., Martino, S., Ball, S. A., & Carroll, K. M. (2011). Mutual influence in therapist competence and adherence to motivational enhancement therapy. *Drug and Alcohol Dependence, 115*, 229–236.
- Kendall, P. C. (1994). Treating anxiety disorders in children: Results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology, 62*, 100–110.
- Kendall, P. C., & Hedtke, K. (2006). *Cognitive-behavioral therapy for anxious children: Therapist manual* (3rd ed.). Ardmore, PA: Workbook Publishing.
- Kendall, P. C., Hudson, J. L., Gosch, E., Flannery-Schroeder, E. F., & Suveg, C. (2008). Cognitive behavioral therapy for anxiety disordered youth: A randomized clinical trial evaluating child and family modalities. *Journal of Consulting and Clinical Psychology, 76*, 282–297.
- Levita, L., Duhne, P. G. S., Girling, C., & Waller, G. (2016). Facets of clinician's anxiety and the delivery of cognitive-behavioral therapy. *Behaviour Research and Therapy, 77*, 157–161.
- McLeod, B. D., Smith, M. M., Southam-Gerow, M. A., Weisz, J. R., & Kendall, P. C. (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.
- McLeod, B. D., Southam-Gerow, M. A., Tully, C. E., Rodriguez, A., & Smith, M. M. (2013). Making a case for treatment integrity as a psychosocial treatment quality indicator for youth mental health care. *Clinical Psychology: Science and Practice, 20*, 14–32.
- McLeod, B. D., & Weisz, J. R. (2010). The Therapy Process Observational Coding System for Child Psychotherapy Strategies Scale. *Journal of Clinical Child & Adolescent Psychology, 39*, 436–443.
- Meyer, J. M., Farrell, N. R., Kemp, J. J., Blakey, S. M., & Deacon, B. J. (2014). Why do clinicians exclude anxious clients from exposure therapy? *Behaviour Research and Therapy, 54*, 49–53.
- Orlinsky, D. E., Rønnestad, M. H., & Willutzki, U. (2004). Fifty years of psychotherapy process-outcome research: Continuity and change. In M. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (5th ed.). New York, NY: Wiley.
- Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science & Practice, 12*, 365–383.
- Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology, 75*, 829–841.
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., ... Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38*, 65–78.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Chicago, IL: Scientific Software International.
- Santa Ana, E. J., Carroll, K. M., Añez, L., Paris, M., Jr., Ball, S. A., Nich, C., ... Martino, S. (2009). Evaluating motivational enhancement therapy adherence and competence among Spanish-speaking therapists. *Drug and Alcohol Dependence, 103*(1-2), 44–51.
- Schoenwald, S. K. (2011). It's a bird, It's a plane, It's... fidelity measurement in the real world. *Clinical Psychology: Science and Practice, 18*, 142–147.
- Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(1), 32–43.
- Schoenwald, S. K., & Hoagwood, K. (2001). Effectiveness, transportability, and dissemination of interventions: What matters when? *Psychiatric Services, 52*, 1190–1197.
- Shaffer, D., Fisher, P., Dulcan, M. K., & Davies, M. (1996). The NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC 2.3): Description, acceptability, prevalence rates, and performance in the MECA study. *Journal of the American Academy of Child and Adolescent Psychiatry, 35*, 865–877.
- Sholomskas, D. E., Syracuse-Siewert, G., Rounsaville, B. J., Ball, S. A., Nuro, K. F., & Carroll, K. M. (2005). 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, 73*, 106–115.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.
- Silverman, W. K., & Albano, A. M. (1996). *The anxiety disorders interview schedule for DSM-IV-child and parent versions*. San Antonio, TX: Graywind Publications.
- Simons, A. D., Padesky, C. A., Montemmarano, J., Lewis, C. C., Murakami, J., Lamb, K., ... Beck, A. T. (2010). Training and dissemination of cognitive behavior therapy for depression in

- adults: A preliminary examination of therapist competence and client outcomes. *Journal of Consulting and Clinical Psychology*, 78, 751–756.
- Southam-Gerow, M. A., Marder, A. M., & Austin, A. A. (2008). Dissemination of evidence-based manualized treatments for children and families in practice settings. In R. G. Steele, D. T. Elkin, & M. C. Roberts (Eds.), *Handbook of evidence-based therapies for children and adolescents: Bridging science and practice* (pp. 447–469). New York, NY: Springer Science + Business Media.
- Southam-Gerow, M. A., & McLeod, B. D. (2013). Advances in applying treatment integrity research for dissemination and implementation science: Introduction to special issue. *Clinical Psychology: Science and Practice*, 20, 1–13.
- Southam-Gerow, M. A., McLeod, B. D., Arnold, C. A., Rodriguez, A., Cox, J. R., Reise, S. P., ... Kendall, P. C. (2016). Initial development of a treatment adherence measure for cognitive-behavioral therapy for child anxiety. *Psychological Assessment*, 28(1), 70–80.
- Southam-Gerow, M. A., Rodriguez, A., Chorpita, B. F., & Daleiden, E. (2012). Dissemination and implementation of evidence based treatments for youth: Challenges and recommendations. *Professional Psychology: Research and Practice*, 43, 527–534.
- Southam-Gerow, M. A., Weisz, J. R., Chu, B. C., McLeod, B. D., Gordis, E., & Connor-Smith (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 and Adolescent Psychiatry*, 49(10), 1043–1052.
- Southam-Gerow, M. A., Weisz, J. R., & Kendall, P. C. (2003). Childhood anxiety disorders in research and service clinics: Preliminary examination of differences and similarities. *Journal of Clinical Child and Adolescent Psychology*, 32, 375–385.
- Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology*, 47A, 631–650.
- Waller, G., & Turner, H. (2016). Therapist drift redux: Why well-meaning clinicians fail to deliver evidence-based therapy, and how to get back on track. *Behaviour Therapy & Research*, 77, 129–137.
- Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, 61, 620–630.
- Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 78(2), 200–211.
- Weersing, V. R., Weisz, J. R., & Donenberg, G. R. (2002). Development of the Therapy Procedures Checklist: A therapist-report measure of technique use in child and adolescent treatment. *Journal of Clinical Child Psychology*, 31, 168–180.
- Weisz, J. R., Jensen, A. L., & McLeod, B. D. (2005). Development and dissemination of child and adolescent psychotherapies: Milestones, methods, and a new deployment-focused model. In E. D. Hibbs & P.S. Peter (Eds.), *Psychosocial treatments for child and adolescent disorders: Empirically based strategies for clinical practice* (2nd ed., pp. 9–39). Washington, DC: American Psychological Association.
- Weisz, J. R., Kuppens, S., Eckshtain, D., Ugueto, A. M., Hawley, K. M., & Jensen-Doss, A. (2013). Performance of evidence-based youth psychotherapies compared with usual clinical care: A multilevel meta-analysis. *JAMA Psychiatry*, 70(7), 750–761.
- Weisz, J. R., Southam-Gerow, M. A., Gordis, E. B., Connor-Smith, J. K., Chu, B. C., Langer, D. A., ... Weiss, B. (2009). 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*, 77(3), 383–396.
- Wood, J. J., Piacentini, J. C., Southam-Gerow, M., Chu, B. C., & Sigman, M. (2006). Family cognitive behavioral therapy for child anxiety disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 43, 314–321.

RECEIVED: March 2, 2016

ACCEPTED: July 27, 2016

AVAILABLE ONLINE: 6 August 2016