HHS Public Access

Author manuscript

J Clin Child Adolesc Psychol. Author manuscript; available in PMC 2020 January 01.

Published in final edited form as:

J Clin Child Adolesc Psychol. 2019 ; 48(SUP1): S234–S246. doi:10.1080/15374416.2017.1381914.

Benchmarking Treatment Adherence and Therapist Competence in Individual Cognitive-Behavioral Treatment for Youth Anxiety Disorders

Bryce D. McLeod[1], **Michael A. Southam-Gerow**[1], **Amanda Jensen-Doss**[2], **Aaron Hogue**[3], **Philip C. Kendall**[4], and **John R. Weisz**[5]

[1]Virginia Commonwealth University, Department of Psychology, PO Box 842018, Richmond, VA 23284- 2018.

[2]University of Miami, Department of Psychology, PO Box 248185, Coral Gables, FL 33124-0751.

[3]The National Center on Addiction and Substance Abuse, 633 Third Avenue, 19th Floor, New York, NY 10017.

[4]Temple University, Department of Psychology, 1701 North 13th Street, Weiss Hall, Philadelphia, Pennsylvania 19122.

[5]Harvard University, Department of Psychology, 1030 William James Hall, 33 Kirkland Street, Cambridge, MA 02138.

Abstract

Objective: Evidence-based treatments (EBTs) for youthare typically developed and established through studies in research settingsdesigned to ensure treatment integrity—i.e., protocol adherence and competence by therapists. An important question for implementation science is how well integrity is maintained when these EBTs are delivered in community settings.The present study investigated whetherthe integrity achieved by therapists in community settings achieved a benchmark set by therapists in a research setting when they delivered the same EBT—an individual cognitive-behavioral treatment (ICBT)for youth anxiety.

Method: Therapists (*N* = 29; 68.97% White; 13.79% male) provided ICBT to 68 youths (M age = 10.60 years, SD = 2.03; 82.35% White; 52.94% male) diagnosed with a principal anxiety disorder in research or community settings. Training and supervision protocols were the same across settings. Two independent teams of trained coders rated 744 sessionsusing observational instruments designed to assess ICBT adherence and competence.

Results: Both adherence and competence were higher in the research setting. Group differences in competence were consistent across treatment, but differences in adherence were most pronounced when treatment shifted to exposure, widely-viewed as the most critical component of ICBT.

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. bmcleod@vcu.edu, mzasouthamger@vcu.edu, ajensendoss@miami.edu, ahogue@centeronaddiction.org, pkendall@temple.edu, john_weisz@harvard.edu

Conclusions: When using the benchmarks from the research setting, therapists from the community settings fell short for indices of adherence and competence. However, given differences between therapists and clients as well as the fact that treatment outcomes were similar across settings, our findings raise questions about whether it is appropriate to use treatment integrity benchmarks from research settings for community.

Keywords

Benchmarking; treatment adherence; competence; cognitive-behavioral treatment; youth anxiety

As a discipline, we aspire to deliver optimal care for youth in community-based mental health settings (hereafter called *community settings*). There is growing pressure for research and community stakeholders to work together to address the public health need to provide high-quality behavioral health services by implementing evidence-based treatments (EBTs) in community settings (Weisz & Kazdin, 2010). One challenge to this effort has been that EBTs were mainly developed in university-based research settings (hereafter called *research settings*), under conditions that differ from those in community settings. So, a goal of implementation research is to understand what happens when EBTs are moved from research to community settings.

Optimizing treatment integrity may be a critical step in achieving the goal of implementing EBTs in community settings.By treatment integrity, we mean the extent to which the elements of a treatment are delivered according to the treatment model (Allen, Linnan, & Emmons, 2012; Schoenwald et al., 2011). Integrity is a multi-faceted construct (Allen et al., 2012), and two of its components are relevant to the current paper (Schoenwald et al., 2011): adherence and competence. Adherence captures the extent of use of interventions prescribed by the treatment protocol, and competence reflects the skill and responsiveness demonstrated by the therapist when delivering the interventions from the protocol. These integrity components may also reflect the degree to which training and supervision impact therapist behavior (Proctor et al., 2011; Schoenwald et al., 2011), and thus predict EBT effectiveness (e.g., Hogue et al., 2008; Robbins et al., 2011).

Successful treatment is thought to depend, in part, on delivering an EBT with integrity (Henggeler, Sheidow, Cunningham, Donohue, & Ford, 2008; Hogue et al., 2008). However, efforts to transport EBTs to community settings have shown that it is difficult to establish and maintain treatment integrity (McHugh & Barlow, 2010; Smith et al., 2016). It is possible thatdifferences between research and community (see Ehrenreich-May et al., 2011; Southam-Gerow, Chorpita, Miller, & Gleacher, 2008)settings may influence treatment integrity. Therapists in research settings may have more specialized training backgrounds consistent with specific EBTs (Bearman et al., 2013), compared to therapists in community settings who have more varied training backgrounds (Santa Ana et al., 2009), which could translate into higher treatment integrity with particular treatments in research settings. Also, youth presenting for treatment in community settings are often more demographically and clinically diverse (Ehrenreich-May et al., 2011; Southam-Gerow et al., 2008). This diversitycould lead therapists to depart from an EBTto meet the clinical needs of the youth, thereby loweringtreatment adherence to the protocol (Smith et al., 2016). It thus is possible

that some of the differences between research and community settings may lead a therapist to deliver a specific EBTdifferently than it is delivered in research settings (Beidas & Kendall, 2010).

Although few studies have directly tested whether treatment integrity varies across settings, some evidence from effectiveness trials suggest that therapists in community settings may deliver relatively low doses of the EBT being tested (Smith et al., 2016;Weisz et al., 2009). Smith et al. (2016), for example, evaluated whether the delivery of the same individual cognitive-behavioral treatment (ICBT) program for youth anxiety differed across research and community settings. The sample was comprised of 89 youth treated by therapists who delivered (a) ICBT in a research setting, (b) ICBT in a community setting, or (c) usual clinical care. The therapists trained and supervised in ICBT for youth anxiety delivered different patterns of cognitive-behavioral interventions over the course of treatment across research and community settings, with therapists in the community settings delivering a lower dose of cognitive-behavioral interventions. However, this study did not use an instrument designed specifically to evaluate adherence to or competence in the delivery of ICBT. Rather, the instrument was designed to assess the delivery of cognitive-behavioral interventions for a wide range of youth emotional and behavioral problems. Because this instrument was not designed for ICBT for youth anxiety, it remains unclear if therapists across research and community settings differ in quantity and quality of interventions prescribed by the ICBT protocol.

Although it is sometimes assumed that treatment adherence and competence for a specific EBT protocol may differ across research and community settings, to our knowledge this question has not been tested. A benchmarking approach provides a possible means of addressing this question. Benchmarking studies evaluate whether therapist performance in community settings approximate the standards achieved by therapists in efficacy trials (Hunsley & Lee, 2007; Spilka & Dobson, 2015). To date, benchmarking studies have primarily focused upon treatment outcomes observed in community settings (e.g., Weersing & Weisz, 2002). However, benchmarking can also be used to study adherence and competence (McLeod, Southam-Gerow, Tully, Rodriguez, & Smith, 2013), when two groups of therapists are tasked with delivering the same treatment program.

The present study evaluated whether therapists in an effectiveness trial conducted in community settings attained treatment integrity at the level of the benchmarkestablished by therapists delivering the same treatment program in a research setting. Our approach to benchmarking analyses was guided by the steps outlined by Weersing (2005): (1) Identify the target problem population and treatment; (2) Identify a gold-standard research benchmark; (c) Assess the outcome(s) in the community setting; and (4) Compare findings from the community setting to the benchmark and explore potential reasons for any differences.

We chose anxiety disorders as the target problem and an efficacious ICBT program, the *Coping Cat*(Kendall & Hedtke, 2006), as the treatment. The *Coping Cat* program is a good candidate for benchmarking as it is comprised of two distinct phases (skill-building, exposure; Kendall & Hedtke, 2006). Exposure is considered to be more difficult to deliver

than skill-building interventions (Balkhi, Reid, Guzick, Geffken, & McNamara, 2016), so it is possible to evaluate whether treatment integrity varies across these two phases. We selected two randomized controlled trials (RCTs), one conducted in a research setting (Kendall et al., 2008) and the other in community settings (Southam-Gerow et al., 2010) that both employed the *Coping Cat* program. Different training and supervision techniques can impact treatment integrity (Sholomskas et al., 2005), so we sought to standardize these procedures across the trials. Thus, both trials used the sameprocedures to train and supervise the therapists in delivering the *Coping Cat* program.To establish benchmark integrity levels, we used two observational instruments that have shown promising scorereliability and validity in research and community samples.

We tested two hypotheses. First, we hypothesized that therapists delivering ICBT in research settings would evidence higher adherence scores than therapists delivering ICBT in community settings. Second, we hypothesized that the competence scores would be higher for therapists delivering ICBT in research settings than for therapists delivering ICBT in community settings. In short, we anticipated that therapists in the community setting would not achieve the benchmark levels of the research setting therapists. We also examined two exploratory questions. First, we evaluated whether adherence and competence changed over time and whether the two settings differed in pattern of change, though we did not formulate specific hypotheses given the contradictory findings regarding change in treatment delivery over time (e.g., Boswell et al., 2013; Henggeler et al., 2008).Second, we examined whether the dose and trajectory of adherence and competence changed over the skill-building and exposure phases of ICBT. Exposure is considered a difficult therapeutic intervention that requires intensive training (Balkhi et al., 2016), so it is possible that adherence and competence may vary across these two phases of ICBT.

Method

Participants

Treatment data were collected from 29 therapists (68.97% White; 13.79% male) who provided ICBT to 68 youth (*M* age = 10.60, *SD* = 2.03; range 7–15 years; 82.35% White; 52.94% male) with a principal anxiety disorder who participated in one of two RCTs, one RCT conducted in a research setting and one RCT conducted in community settings (see Kendall et al., 2008 and Southam-Gerow et al., 2010 for details). Treatment data included recorded sessions collected in each RCT. To be included in this study, youth had to: (a) have at least two audible recorded sessions, and (b) have received ICBT from a single therapist. This study was institutional review board approved. Parents provided written informed consent, and youth gave written or verbal assent. See Tables 1 and 2 for descriptive information about youth and therapist participants.

Research setting.—Kendall et al. (2008) compared the efficacy of ICBT (*n* = 55), family-CBT (*n* = 56), and a family-based education/support/attention control group (*n* = 50). Only ICBT was used in this study. The 51 youth participants (*M* age = 10.36 years, *SD* = 1.90; 86.28% White; 60.78% male) from the ICBT group included in this study received ICBT at a university-based research clinic that specialized in the treatment of anxiety disorders.

Therapists (n = 16; 12.50% male) were mostly White (81.25%); some were Latino and Asian/Pacific Islander (both 6.25%), and 6.25% did not report. 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.

Community settings.—The Youth Anxiety Study (YAS; Southam-Gerow et al., 2010) compared the effectiveness of ICBT (n = 24) to usual care (n = 24). Only ICBT (YAS-ICBT) was used in this study. The 17 youth participants in YAS-ICBT (M age = 11.32 years, SD = 2.32; 41.18% White; 29.41% male) included in this study received treatment at community-based mental health clinics across Los Angeles county. All therapists were clinic employees who volunteered to participate in the study and were randomly assigned to groups. Therapists assigned to YAS-ICBT (n = 13; 15.38% male) were 53.86% White, 15.38% Latino, 15.38% Asian/Pacific Islander, and 15.38% mixed/other. Professional composition of the therapists was 30.76% social workers, 38.48% psychology (23.14% masters level, 15.34% doctoral level), and 30.76% reported “other” degree. YAS-ICBT therapists endorsed a variety of theoretical orientations including 38.48% psychodynamic, 30.76% cognitive-behavioral, 15.38% family systems, and 15.38% “other”. The therapists reported an average of 3.92 years (SD = 1.934; range 1 to 7) of training and an average of 9.00 years (SD = 11.65; range 0 to 35) of post-training clinical experience.At post-treatment, 66.70% of youths in the original study no longer met diagnostic criteria.

Individual Cognitive Behavioral Treatment

Therapists in ICBT and YAS-ICBT delivered *Coping Cat*, an ICBT program for youth diagnosed with anxiety disorders (Kendall & Hedtke, 2006). The program consists of 16 sessions; 14 sessions are conducted individually with the youth and two sessions are conducted with the parents. The first half focuses on anxiety management skills training (e.g., relaxation, problem-solving) and the second half emphasizes exposure tasks (i.e., exposure start at session nine). Homework is regularly assigned to the youth. In both studies, gold-standard quality control methods used in RCTs were employed to help establish and maintain treatment integrity, which included a treatment protocol, a training workshop, and weekly supervision with an expert in CBT for youth anxiety (Sholomskas et al., 2005). In both original RCTs, adherence to *Coping Cat* wasmeasured with the Coping Cat Brief Adherence Scale (seeKendall, 1994), which uses a checklist format (presence/absence) to determine if core ICBT interventions were delivered. Based on the scale, therapists in both studies showed more than 90.00% adherence. (See Kendall et al., 2008 and Southam-Gerow et al., 2010 for details).

Adherence and Competence Instruments

Cognitive Behavioral Treatment for Anxiety in Youth Adherence Scale—(CBAY-A; Southam-Gerow et al., 2016) is a 22-item instrument that assesses three areas: (a) *Standard* four items that represent interventions found in most CBT sessions (i.e., Agenda Setting, Homework Review, Homework Assigned, Rapport Building), (b) *Model* 12 items that assess model-specific content (i.e., Psychoeducation Anxiety, Emotion Education, Fear Ladder, Relaxation, Cognitive Anxiety, Problem Solving, Self-Reward, Coping Plan,

Exposure Preparation, Exposure, Exposure Debrief, Maintenance), and (c) *Delivery* six items that measure how model items are delivered (i.e., Didactic Teaching, Collaborative Teaching, Modeling, Rehearsal, Coaching, Self-Disclosure).Coders watch entire sessions and rate each item on a 7-point extensiveness scale: 1 = *not at all*, 3 = *somewhat*, 5 = *considerably*, 7 = *extensively*. CBAY-A item scores have demonstrated evidence of construct validity (Southam-Gerow et al., 2016). For the benchmarking analyses, we used two subscales that map onto the two phases in ICBT: Skills Phase (Psychoeducation Anxiety, Emotion Education, Fear Ladder, Relaxation, Cognitive Anxiety, Problem Solving, Self-Reward, Coping Plan), and Exposure Phase (Coping Plan, Exposure Preparation, Exposure, Exposure Debrief). Subscale scores were produced by taking the highest scoring item from the subscale for each session (see Southam-Gerow et al., 2016; Smith et al., 2016). For example, the Exposure subscale was scored by taking the highest score from the Coping Plan, Exposure Preparation, Exposure, or Exposure Debrief item for each session whenever any of those items were scored over the course of treatment.It is important to note that although the Coping Cat program is designed to start with skill building and advance to exposures, not all cases in the sample progressed through the phases in the same way. For example, some cases started exposures before session 9, which is the typical starting point for exposures in the Coping Cat program. This means that skill building and exposure interventions could be scored, for some cases, over the course of treatment. For the present sample, the inter-rater reliability of the Skills and Exposure subscales fell in the “excellent” range (Cicchetti, 1994), ICC(2,2), and was 0.78 and 0.88, respectively.

Cognitive Behavioral Treatment for Anxiety in Youth Competence Scale— (CBAY-C; McLeod et al., 2016) is a 25-item observational instrument that parallels the content of the CBAY-A. The CBAY-C assesses the same three areas as the CBAY-A (Standard, Model, Delivery) along with two global items that assess level of responsiveness and skillfulness of CBT delivery across a session. In making competence ratings, coders are asked to watch an entire session and make ratings for each item that is observed on a 7-point Likert-type competence scale with the following anchors: 1 = *very poor*; 3 = *acceptable*; 5 = *good*; 7 = *excellent*. Items that are not observed during a session are not scored. The CBAY-C Scale, Subscale, and item scores have demonstrated evidence of construct validity (McLeod et al., 2016). However, the CBAY-C Subscale scores were highly correlated in this sample ($r = .86$, $p < .001$), so we used the Total scale for analyses (Psychoeducation Anxiety, Emotion Education, Fear Ladder, Relaxation, Cognitive Anxiety, Problem Solving, Self-Reward, Coping Plan, Exposure Preparation, Exposure, Exposure Debrief). The Total scale score was produced by taking the highest scoring item from the scale for each session (see McLeod et al., 2016; Smith et al., 2016). For the present sample, the CBAY-C scale inter-rater reliability, ICC(2,2), was .77, which falls in the “excellent” range (Cicchetti, 1994). The correlation between the CBAY-A Skills and Exposure subscales and the CBAY-C scale was .43 ($p < .01$) and .33 ($p < .01$), respectively.

Coding and Session Sampling Procedures

Six doctoral student coders comprised the coding teams; two were part of the CBAY-A team and four were part of the CBAY-C team. Training was conducted by the first and second authors and progressed through the same steps for each instrument. 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, coder independently coded 32 recordings and reliability was assessed against master codes produced by the first and second authors. To be certified for independent coding, each coder had to demonstrate "good" reliability on each item (ICC > .59; Cicchetti, 1994). Following certification, regular meetings were held to help prevent coder drift. Sessions were randomly assigned to coders who were naïve to study hypotheses. 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, because these sessions may contain intake or termination content. Existing recordings were not rated if (a) audible content was shorter than 15 minutes (n = 25), or (b) less than 75% of the dialogue was in English (n = 3).

Assessments Collected in the Original RCTs

The original RCTs collected diagnostic and symptom instruments that were used in this study for group comparisons and control purposes. Kendall et al. (2008) 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. 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. Both studies collected the *Child Behavior Checklist* (CBCL; Achenbach, 1991) to assess symptoms across three broad-band scales (e.g., Internalizing) and eight narrow-band sub-scales (e.g., Somatic Complaints). In the current investigation, t-scores on four CBCL scales were used: Total, Internalizing (broad-band), Externalizing (broad-band), and Anxious-Depressed (narrow-band).

Data Analysis Plan

Analyses of group differences in adherence and competence were conducted using multilevel modeling (Raudenbush & Bryk, 2002) using HLM 7.01 (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011) to account for the nesting of youth in therapists and repeated measures in youth. For the CBAY-A subscale scores, change over time was modeled via a piecewise linear model using an increment/decrement model (Raudenbush & Bryk, 2002) that included a base rate slope parameter estimating change over all sessions and an increment/decrement parameter that estimated change in slope after the exposure phase of treatment started. A dummy coded variable for phase was also entered at level 1 to model mean differences between the two phases of ICBT: skills building and exposure. For analysis of the CBAY-A Skills subscale, slopes were centered at the first session of the skills phase of treatment; the CBAY-A Exposure subscale models were centered at the first exposure session for each case. Group differences were evaluated by dummy coding group membership and entering it at a youth-level (level 2). Below is an example of a 3-level odel that was fit to the CBAY-A Skills subscale scores, where TIMEWEEK is the baseline slope, EXPCHG is the increment/decrement slope, EXPOSURE is the dummy coded phase variable (exposure = 1, skills training = 0), and ICBT is the group difference variable (ICBT = 1, YAS-IBCT = 0):

$$\text{L1: CBAYASKILLS}_{ijk} = \pi_{0jk} + \pi_{1jk}^{*}(TIMEWEEK_{ijk}) + \pi_{2jk}^{*}(EXPCHG_{ijk}) + \pi_{3jk}^{*}(EXPOSURE_{ijk}) + e_{ijk}$$

$$\text{L2: } \pi_{0jk} = \beta_{00k} + \beta_{01k}^{*}(ICBT_{jk}) + r_{0jk}\pi_{1jk} = \beta_{10k} + \beta_{11k}^{*}(ICBT_{jk}) + r_{1jk}\pi_{2jk} = \beta_{20k} + \beta_{21k}^{*}(ICBT_{jk}) +$$

$$r_{2jk}\pi_{3ik} = \beta_{30k} + \beta_{31k}^{*}(ICBT_{jk}) + r_{3k}$$

$$\text{L3: } \beta_{00k} = \gamma_{000} + u_{00k}\beta_{01k} = \gamma_{010}\beta_{10k} = \gamma_{100} + u_{10k}\beta_{11k} = \gamma_{110}\beta_{20k} = \gamma_{200} + u_{20k}\beta_{21k} = \gamma_{210}\beta_{30k} = \gamma_{300} +$$

$$u_{30k}\beta_{31k} = \gamma_{310}$$

Although examination of fit statistics supported the use of these models for the CBAY-A subscales, initial examination of the CBAY-CTotal scale indicated very little change over time in the scores, and a comparison of fit statistics (AIC and BIC) for the model described above and an intercept-only model without change over time indicated the intercept only model fit the data best. So, group differences in the CBAY-C were examined using an intercept-only model where repeated measures were used to estimate a mean level of competence for each case.

Cohen's *d* effect sizes were calculated. For parameters representing mean differences, effect sizes were computed by dividing the group difference parameter estimate by the raw data SDs. For slope parameter, effect sizes were computed following Feingold's (2009) recommendation to multiply the parameter by the length of treatment (or the length of exposure in the case of the increment/decrement slope) 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 and the average length of exposure of eight weeks. All effect size calculations represent Cohen's (1988) *d*, for which the following interpretation parameters have been suggested: .20 small, .50 medium, and .80 large.

Finally, we examined whether the group differences remained significant when controlling for youth demographic and clinical characteristics that differed between the groups. Based on missing data analyses described below, 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 22.00. Control variables were entered simultaneously into each subscale model, and continuous control variables were grand mean centered.

Results

Preliminary Analyses

We conducted sample bias analyses to determine if the 68 youthand 29 therapistsin this study differed from the participants in the parent studies on demographiccharacteristics, clinical characteristics, and baseline therapist training characteristics (see Tables 1 and 2). We found a lower proportion of African-American youth (0% vs 16.67%) and a higher

proportion of White youth (41.18% vs. 29.17%) in our YAS-ICBT sample compared to the parent study, $X^2(3, N = 24) = 11.53$, $p = .009$).

As detailed in Table 1, we examined differences between groups (ICBT, YAS-ICBT) on several youth demographic, baseline, and clinical characteristics. Groups differed on sex, race/ethnicity, level of externalizing symptomatology, level of anxiety and depressive symptomatology, principal anxiety disorder, family income level, average length of a treatment session, weeks in treatment, and total minutes spent in treatment (i.e., number of sessions multiplied by the average session length in minutes). As reported in Table 2, groups did not differ on therapist sex or race/ethnicity, but did differ on therapist professional training.

We next examined missing data. Of the 1098 sessions held, 744 (67.75%) were rated (65.50% ICBT (n = 532), 74.10% YAS-ICBT (n = 212)). There was no significant difference between groups in terms of the percent of sessions coded, $t(66) = 1.85$, p = .069, nor was there a difference in the percent of sessions coded from the first and second half of treatment (first half = 67.64%; second half 67.88%; $t(67) = 0.07$, p = .95).We also evaluated whether any of the child demographic or clinical characteristics related to the percent of session coded per case and did not find that any of the characteristics were related (all ps > .20). As multilevel models can accommodate varying amounts of repeated measures data across participants, missing data analyses primarily focused on patterns of missingness in the youth-level control variables. Residuals files from the multilevel models described above were used to generate youth-level intercept and slope estimates that were included in the missing data check to account for possible relations between missing control variable data, adherence, and competence. Rates of missing data were 8.80% (race/ethnicity, income) or less across control variables. These data were missing completely at random (Little's MCAR test $X^2 = 108.40$, $DF = 92$, $p = .12$).

Treatment Integrity Benchmarks

CBAY-A Skills subscale.—Prior to examining group differences, we examined whether variability in scores on the CBAY-A Skills subscale was located at the youth or therapist level. Fully 94.00% of the variability in intercept (i.e., scores on the Skills subscale inthe first session), 67.00% of the variability in change over time, 39.00% of the variability in slope change during exposure, and 53.00% of the mean differences in scores on the Skills subscale between the skills and exposure phases were located at the therapist level. Model results examining group differences are presented in Table 3. Relative to YAS-ICBT, ICBT had significantly higher scores on the Skills subscaleat the firstsession ($\gamma_{010} = 1.01$, p = .003, d = .57). Both groups demonstrated a decline in scores on the CBAY-A Skillssubscale after the initiation of exposure; there were no group differences in these patterns. Model plots to illustrate these patterns are presented in Figure 1.

CBAY-A Exposure subscale.—For the CBAY-A Exposure subscale, 45.00% of the variability in intercept (i.e., scores on the Exposure subscale in the first exposure session), 36.00% of the variance in slope, 14.00% of the change in slope following exposure, and 68.00% of the mean differences between the skills and exposure phases were located at the

therapist level. Model results examining group differences are presented in Table 3.At the first exposure session, ICBT had significantly higher scores on the CBAY-A Exposure subscalethan YAS-IBCT (γ_{010} = 1.27, p < .001, d = 1.17). Over treatment, there was also a significant group difference in the CBAY-A Exposure slope (γ_{110} = .14, p < .001, d = 2.95); the ICBT group showed increases in exposure adherence over time, whereas the YAS-ICBTscores remained flat. Both groups demonstrated an increase in scores on the CBAY-A Exposure subscale following the beginning of exposure, but that increase was larger for ICBT (γ_{310} = .88, p = .043, d = .81). After the beginning of the exposure phase, ICBT showed a steeper decline in the slope than YAS-ICBT (γ_{110} = −.20, p < .001, d = −1.47); this indicated that the YAS-ICBT group continued to show flat scores over the exposure phases, whereas scores for the ICBT group decreased over the exposure phase.Model results are plotted in Figure 1.

CBAY-C Total scale.—Fully 99.00% of the variance in scores on the CBAY-C Total scale was located at the therapist level. ICBT had significantly higher competence than YAS-ICBT (γ_{010} = 1.53, p < .001, d = 1.13).

Ruling Out Alternative Interpretations

We examined whether the treatment integrity findings held when including characteristics that differed across groups in the model: Youth sex, race/ethnicity, CBCL Externalizing, CBCL Anxious-Depressed, principalanxiety diagnosis, and family income. Given that weeks in treatment, average session length, and total minutes of treatment all represented treatment dose, we controlled for total minutes of treatment, as that encompassed both length of treatment and length of sessions. Control variables were entered simultaneously into each scale and subscale model, using multiple imputation to account for missing data across groups. Results are presented in Table 3. For the CBAY-A Skills subscale scores, the group difference in scores at the first session remained significant when controlling for youth characteristics and the group difference in the mean change after the initiation of exposure remained non-significant. However, two new group differences emerged, suggesting a suppressor effect of the youth variables. In these analyses, YAS-ICBT increased in scores on the Skills subscale over the skills phase of treatment, whereas the ICBT group remained flat. After the initiation of the exposure phase, the YAS-ICBT demonstrated a mean drop in scores on the Skills subscale, followed by a subsequent additional small decline in scores on the Skills subscale. ICBT also had a drop in scores on the Skills subscale after the beginning of the exposure phase, but then remained relatively flat. For the CBAY-A Exposure subscale, none of the group differences remained significant. However, the effect sizes associated with those effects remained medium to large, suggesting that these changes in p value were likely attributable to additional variables being in the model rather than these controlvariables explaining away the relations.For the CBAY-CTotal scale, the group difference remained significant when controlling for youth characteristics. Figure 1 depicts the CBAY-A Skills and Exposure models, controlling for youth and treatment characteristics.

To better understand the changes in the models after including the control variables, post hoc analyses that examined each control variable one at a time were conducted. Changes between the original CBAY-A Skills model and the model including the control variables

seemed attributable to group differences in a principal diagnosis of specific phobia, as the group difference in the overall slope for scores on the CBAY-A Skills subscale became significant in models controlling for that variable. However, group differences in the mean score and change in slope after the onset of exposure were not significant in any of the models with individual control variables. To further probe the findings related to CBAY-A Skills subscale, we removed the group predictors from the models to evaluate how a principal diagnosis of specific phobia related to scores on the CBAY-A Skills subscale. Specific phobia was related to the drop in scores on the CBAY-A Skills subscale following the onset of exposure. Specifically, smaller drops were observed for cases with a principal diagnosis of specific phobia. However, that variable was not related to the slopes, which was where it appeared to impact the model results. Overall, these findings suggest that the suppressor effects for the CBAY-A Skills subscale scores may be a function of multiple variables in combination and not due to any specific variable included in the model.

We also performed post hoc analyses to investigate the changes in the CBAY-A Exposure subscale scores. We found that the group differences in overall slope and change in slope after exposure became non-significant in the model controlling for treatment dose. The group difference in the jump in exposure adherence after the initiation of exposure was no longer significant in the models controlling for income and a principal diagnosis of generalized anxiety disorder. Group differences in levels of adherence during the first exposure session remained significant in all of the individual models. Overall, it appears that when a few specific control variables were entered one at a time into the model some changes in significance were observed. However, though the significance levels were impacted, the effect size values continued to be medium to large, suggesting that the control variables did not impact the magnitude of the findings, just the significance.

Discussion

The purpose of this study was to examine whether therapists in a community setting achieved the benchmark level of adherence and competence in the delivery of ICBT for youth anxiety produced by therapists in a research setting.Consistent with our hypotheses, the community therapists did not achieve the benchmark scores of adherence or competence of the research clinic therapists, suggesting that the quantity and quality of ICBT delivery differed across settings. We also found that treatment adherence systematically changed over treatment, with adherence during the exposure phase increasing more steeply in the research setting. In contrast, competence did not change over treatment in either setting. Differences in adherence and competence observed between settingslargely held when differences in youth characteristics between the groups were included in the model. Together, our findings have implications for efforts to transport EBTs to community settings.

Our findings are in accord with previous research that has suggested that when therapists in community settings deliver EBTs they do so at a lower level of extensiveness (i.e., a lower dose)than therapists in research settings (Weisz et al., 2009; Wood et al., 2006).When we entered relevant variables into the model that were available (e.g., youth race/ethnicity, comorbid symptoms), the group differences remained (i.e., effect sizes remained medium to large), suggesting that the factors assessed in our study did not account for the observed

differences. However, our measurement model did not contain many possibly relevant variables, especially related to the therapists.

Beyond overall differences observed across settings in adherence and competence, we also found important differences in adherence within and between groups over the course of ICBT for youth anxiety.As one example, the largest gap between research and community therapist adherence was observed around the ninth session, which is when the exposure tasks typically start (exposure being a component that significantly improves the trajectory of improvement, Peris et al, 2015). At that time, therapists in the research setting delivered a significantly higher dose of exposure interventions than therapists in the community settings (d = .81).We also found that the trajectory of both the skill-building and exposure adherence scores varied over treatment.Notably, the trajectory of the scores on the Skills and Exposure subscales differed across the two ICBT phases, suggesting that there may be value in breaking down the measurement of adherence into phases for this treatment model.One possible reason for the finding is that the skill building phase may be easier to transport with integrity, as it takes place largely in an office setting where the exposure phase may be more challenging, as it often involves arranging out-of-office activities that place the youth in anxiety-provoking situations. Alternatively, there is some evidence that therapists are wary of delivering exposure interventions (Becker, Zeifert, & Anderson, 2004s; Deacon et al., 2013), suggesting that integrity to exposure could be dampened based on therapist attitudes toward the intervention. Some have hypothesized that exposure is important to positive outcomes for youth anxiety (Peris et al, 2015); future studies should thus examine the importance of adherence to the exposure phase for treatment outcome.

Together, our findings suggest that the adherence may vary from session to session (see Dunn et al., 2016).This finding contributes to a small body of research indicating that treatment adherence may change over an episode of care (e.g., Chiapa et al., 2015; Robbins et al., 2011). If so, there is a need to assess adherence multiple times over the course of a multi-phase treatment to produce an accurate estimate (Dennhag, Gibbons, Barber, Gallop, & Crits-Christoph, 2012).

Unlike the adherence scores, competence scores did not vary over treatment.This suggests that competence levels may remain stable within a case, consistent with previous findings (e.g., Crits-Christoph et al., 1998).The lack of variability over treatment and the fact that competence scores from the skill-building and exposure phases of ICBT were highly correlated suggest that efforts to generate separate competence scores for the two phases of ICBT may not bear fruit (McLeod et al., 2016).That said, our findings indicate that competence scores may not be influenced by factors that occur within specific cases.Indeed, 99.00% of the variance was at the therapist level suggesting that variations in competence scores were therapist-level factors.Future research should investigate if competence scores vary across cases, as such findings could have implications for efforts to train therapists.

Acautionary note is warranted about interpreting our findings related to differences between the research and community settings.Our findings provide information about the similarities and differences in the delivery of ICBT for youth anxiety across settings.However, our findings do not identify what adherence and competence scores are needed to achieve

optimal treatment outcomes in either setting (i.e., anintegrity/outcomered line; Shaw, 1984).Thus, our findings raise questionsabout whether it is appropriate to use the research clinic adherence and competence scores as a benchmark for the delivery of ICBT in community settings. Although these data suggest that therapists in the community settings did not achieve the benchmark, treatment outcomes were similar across the two trials (Kendall et al., 2008; Southam-Gerow et al., 2010). It is conceivable that different levelsof adherence and competence are sufficient depending on setting.If this were the case, benchmarking integrity may be more context-dependent, meaning establishing a universal benchmark or gold-standard may not be a pertinent goal. On the other hand, it is possible that boosting adherence and competence scores in the community settings may lead to even better outcomes. To answer these questions will require further research into the nature of treatment integrity-outcome relations across settings. Thus far, research has failed to find a consistent relation between treatment integrity and treatment outcomes (see Webb, DeRubeis, & Barber, 2010). Clearly, more research is needed to help understand how best to optimize treatment integrity in order to produce favorable treatment outcomes acrossdifferent settings.

Though optimizing treatment integrity is considered critical to the success of treatment (Proctor et al., 2011), little is known about what levels of adherence and competence are needed to achieve this aim in different settings.Our findings suggest that optimization may not mean maximization, since treatment outcomes were similar across both studies. This is consistent with another community-based study that found the program with lower levels of adherence to the sequenceof the protocol achieved better outcomes than the program with higher sequencing adherence scores (Park etal., 2015). Importantly, our findings only speak to quantity and quality of core ICBT interventions as our coding did not capture other non-prescribed interventions delivered by the therapists. It is possible that the community-based therapists had lower adherence and competence scores because they delivered interventions not associated with ICBT to address comorbidities. This type of flexibility in delivering EBTs may be needed in community settings to achieve optimal outcomes (Weisz et al., 2012). So, although comparing community versus research settings provides useful information, it may be most appropriate to identify benchmarks that are appropriate for a particular setting (or context) rather than selectinga single “gold-standard” benchmark to be applied across all settings (McLeod et al., 2013).It might, for example, be possible to establish benchmarks for a specific setting by determining what level of adherence and competence scores are associated with desired treatment outcomes (e.g., below a clinical cutoff at post-treatment).A key future step will be to establish the levels of adherence and competence that produce the best outcomes—and the extent to which these are similar or different across settings.

Future research could also address whether certain factors not measured in our study explain the observed differences in adherence and competence scores. As ICBT for youth anxiety is delivered in the context of a therapeutic relationship, there are potential sources of variability in adherence and competence scores – i.e., client (e.g., symptom severity; Boswell et al., 2013), therapist (e.g., attitudes towards EBTs; Beidas, Edmunds, Marcus, & Kendall, 2012), dyadic (interaction between client and therapist), and error (not specifically categorized; Imel et al., 2014). We assessed for some relevant youth factors. Interestingly, when we

controlled for differences in youthcharacteristics between groups some of the differences widened. This suggests that, at least for skills adherence, research clinic therapists show even higher adherence than community therapistswhen therapistsare treating the same kinds of clients. However, we did not assess for all factors that previously have been found to impact adherence or competence scores (e.g., dyadic factors such as race/ethnicity match between therapist and client; Chapman & Schoenwald, 2011). Moreover, our study design did not permit a thorough assessment of therapist or interaction effects, as there were not enough youth nested within therapists. Few studies have assessed the potential impact of these sources of variance on adherence and competence, especially in the youth treatment field, so this represents an exciting direction for future research.

Additionally, our findings raise some questions for future research to address about the extent to which therapist training and supervision influenced adherence and competence. The research clinic therapists were in an ideal context for maximizing integrity, operating in a clinic that delivered a single treatment program to a focused population under the training and supervision of world experts in the treatment model, including the treatment developer. In contrast, the full-time community therapists were working in a context in which only one or two of their20–30 cases were in the study. Though the training and supervision for the ICBT program was comparable for both groups, it is possible that other factors may have influenced the delivery of ICBT. For example, the therapists in the research clinic may have received more foundational training in CBT as part of their graduate training. More foundational training and supervision in ICBT for youth anxiety may boost scores of therapists in the community settings as could more overall experience with the treatment model.The transfer of training literature suggests one key variable in optimizing integrity is the opportunity to perform that skill (e.g., Bearman et al., 2013; Ford & Weissbein, 1997). The therapists in the community settings were relatively inexperienced in the delivery of ICBT for youth anxiety compared to their research counterparts and they had limited opportunity to practice their newly learned skills, seeing one or two cases total in two years among their other cases. Specific to competence, some past work has suggested that there is an initial boost in competence following training (Simons et al., 2010), though the evidence that competence increases with more experience has shown mixed findings (e.g., Crits-Christoph et al., 1998; Simons et al., 2010).Overall, this literature is quite sparse and considerably more work is needed to understand how training and supervision specifically influence performance.

Limitations to the study bear mentioning. First, each therapist saw very few cases, especially in the community setting; thus, it was difficult to tease apart youth and therapist effects. Second, our ability to identify factors that accounted for the observed differences between the research and community settings was likely limited by the small number of youth and therapists and the limited number of variables that were collected in both studies.Third, therapist professional background and training was largely confounded with setting in our sample making it difficult to ascertain if there is indeed a relation between training, adherence, and competence.Fourth, though the type of training in ICBT was similar across settings, the amount of foundational training in CBT may have differed across the samples. Unfortunately, this information was not recorded so we are unable to investigate whether it influenced the delivery of ICBT.Fifth, the two trials were selected for this study because they

utilized the same therapist training and supervision procedures. However, integrity to the training and supervision procedures was not assessed. It is possible that differences between the trials in the quality of training and/or supervision may have influenced the findings.Finally, because no sample can be completely representative of treatment delivery in research and community settingsit will be important to replicate these comparisons across a wider range of settings, treatments, and youth problems.

The limitations should be considered within the context of the study strengths. First, this is one of the first studies to code all available sessions from the same EBT delivered across research and community settings. This afforded the rather unique opportunity to investigate trajectories of adherence and competence over treatment. Second, we used adherence and competence instruments that have initial evidence supporting score reliability and validity. Third, the two parent studies followed the same quality control procedures to deliver the same EBT across settings, which included a treatment protocol, a training workshop, and weekly supervision with an expert in CBT for youth anxiety.

Overall, our findings suggest that community clinic therapists did not achieve the adherence or competence benchmarks established by research clinic therapists. However, our findings raise questions about whether applying benchmarks from research settings to community settings is useful. Though adherence and competence scores differed, post-treatment outcomes were similar across settings. That there were similar outcomes potentially undermines the argument that we should apply integrity benchmarks across settings. Alternatively, it is possible that boosting adherence and competence scores in the community settings might have further improved the treatment outcomes (e.g., certifying therapists in particular approaches). So, although some have recommended benchmarking community clinic performance to research clinic levels (e.g., Weersing, 2005), there are other plausible approaches such as benchmarking to an external but similar comparator (e.g., Stern, Niemann, Wiedemann, & Wenzlaff, 2011) or using internal benchmarks as a method to drive improvement through increasing scores over time (e.g., Pincus, Spaeth-Rublee, & Watkins, 2011). It is possible that differences in contexts, therapists, and clients may necessitate different levels of adherence and competence. Future research will be needed to determine whether establishing benchmarks for community settings based on treatment outcomes may be a better path forward for using benchmarking as a quality improvement approach.

Acknowledgements:

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

References

Achenbach TM (1991). Integrative guide for the 1991 CBCL/4–18, YSR and TRF profiles. Burlington: Department of Psychiatry, University of Vermont.

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

Balkhi AM, Reid AM, Guzick AG, Geffken GR, & McNamara JPH (2016) The Progressive Cascading Model: A scalable model for teaching and mentoring graduate trainees in exposure therapy. Journal of Obsessive-Compulsive and Related Disorder, 9, 36–42

Bearman S, Weisz JR, Chorpita B, Hoagwood K, Ward A, Ugueto A, & 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(6), 518–529. [PubMed: 23525895]

Becker CB, Zayfert C, & Anderson E (2004). A survey of psychologists' attitudes towards and utilization of exposure therapy for PTSD. Behaviour Research and Therapy, 42(3), 277–292. [PubMed: 14975770]

Beidas RS, Edmunds JM, Marcus SC, & Kendall PC (2012). Training and consultation to promote implementation of an empirically supported treatment: A randomized trial. Psychiatric Services, 63(7), 660–665. [PubMed: 22549401]

Beidas R & Kendall PC (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. [PubMed: 20877441]

Boswell JF, Gallagher MW, Sauer-Zavala SE, Bullis J, Gorman JM, Shear MK, … Barlow DH. (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. [PubMed: 23339537]

Chapman JE, & Schoenwald SK (2011). Ethnic similarity, therapist adherence, and long-term multisystemic therapy outcomes. Journal of Emotional and Behavioral Disorders, 19(1), 3–16. [PubMed: 24999302]

Chiapa A, Smith JD, Kim H, Dishion TJ, Shaw DS, & Wilson MN (2015). The trajectory of fidelity in a multiyear trial of the Family Check-Up predicts change in child problem behavior. Journal of Consulting and Clinical Psychology, 83(5), 1006–1011. [PubMed: 26121303]

Cicchetti DV (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.

Crits-Christoph P, Siqueland L, Chittams J, Barber JP, Beck AT, … Woody G(1998). Training in cognitive, supportive-expressive, and drug counseling therapies for cocaine dependence. Journal of Consulting and Clinical Psychology, 66(3), 484–492. [PubMed: 9642886]

Deacon BJ, Farrell NR, Kemp JJ, Dixon LJ, Sy JT, Zhang AR, & McGrath PB, (2013). Assessing therapist reservations about exposure therpay for anxiety disorders: The Therapist Beliefs about Exposure Scale. Journal of Anxiety Disorders, 27(8), 772–780. [PubMed: 23816349]

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

Dunn C, Darnell D, Atkins DC, Hallgren KA, Imel ZE, Bumgardner K, Owens M, & Roy-Byrne P (2016). Within-provider variability in motivational interviewing integrity for three years after MI training: Does time heal? Journal of Substance Abuse Treatment, 65, 74–82. [PubMed: 27016875]

Ehrenreich-May J, Southam-Gerow MA, Hourigan SE, Wright LR, Pincus DB, & Weisz JR (2011). Characteristics of anxious and depressed youth seen in two different clinical contexts. Administration and Policy in Mental Health and Mental Health Services Research, 38(5), 398–411. [PubMed: 21197564]

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. [PubMed: 19271847]

Ford JK & Weissbein DA (1997). Transfer of training: An update review and analysis. Performance Improvement Quarterly, 10, 22–41

Henggeler SW, Sheidow AJ, Cunningham PB, Donohue BC, & Ford JD (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 and Adolescent Psychology, 37, 682–689. [PubMed: 18645758]

Hogue A, Henderson CE, Dauber S, Barajas PC, Fried A, & Liddle HA (2008). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. Journal of Consulting and Clinical Psychology, 76, 544–555. [PubMed: 18665684]

Hunsley J, & Lee CM (2007). Research-informed benchmarks for psychological treatments: Efficacy studies, effectiveness studies, and beyond. Professional Psychology: Research and Practice, 38, 21–33.

Imel ZE, Baldwin SA, Baer JS, Hartzler B, Dunn C, Rosengren DB, & Atkins DC (2014). Evaluating therapist adherence in motivational interviewing by comparing performance with standardized and real patients. Journal of Consulting and Clinical Psychology, 82(3), 472–481. [PubMed: 24588405]

Kendall PC (1994). Treating anxiety disorders in children: Results of a randomized clinical trial. Journal of Consulting and Clinical Psychology, 62, 100–110. [PubMed: 8034812]

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

Kendall PC, Hudson JL, Gosch E, Flannery-Schroeder E, & 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. [PubMed: 18377124]

McHugh RK & Barlow DH (2010). The dissemination and implementation of evidence-based psychological treatments. A review of current efforts. American Psychologist, 65(2), 73–84. [PubMed: 20141263]

McLeod BD, Southam-Gerow MA, Rodriguez A, Quinoy A, Arnold C, Kendall PC, & Weisz JR (2016). Development and initial psychometrics for a therapist competence instrument for CBT for youth anxiety. *Journal of Clinical Child and Adolescent Psychology*. Advance online publication.

McLeod BD, Southam-Gerow MA, Tully CB, Rodriguez A, & Smith MM (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. [PubMed: 23935254]

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

Peris T, Compton S, Kendall PC, Birmaher B, Sherill J, March J, … & Piacentini J (2015). Trajectories of change in youth anxiety during cognitive-behavior therapy. Journal of Consulting and Clinical Psychology, 83, 239–252. [PubMed: 25486372]

Pincus HA, Spaeth-Rublee B, & Watkins KE (2011). The case for measuring quality in mental health and substance abuse care. Health Affairs, 30(4), 730–736. [PubMed: 21471495]

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

Raudenbush SW, & Bryk AS (2002). Hierarchical linear models: Applications and data analysis methods *(2*nd *ed.)* Newbury Park, CA: Sage.

Raudenbush SW, Bryk AS, Cheong YF, Congdon RT, & du Toit M (2011). HLM 7: Hierarchical linear and nonlinear modeling. Chicago, IL: Scientific Software International.

Robbins M, Feaster D, Horigian V, Puccinelli M, Henderson CE, & Szapocznik J (2011). Therapist adherence in Brief Strategic Family Therapy for adolescent drug abusers. Journal of Consulting and Clinical Psychology, 79, 43–53. [PubMed: 21261433]

Santa Ana EJ, Carroll KM, Añez L, Paris M, Ball S, Nich C, Frankforter TL, Suarez-Morales L, Szapocznik J, & Martino S (2009). Evaluating motivational enhancement therapy adherence and competence among Spanish-speaking therapists. Journal of Drug and Alcohol Dependence, 103, 44–51. [PubMed: 19394164]

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

Shaffer D, Fisher P, Dulcan MK, & 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. [PubMed: 8768346]

Shaw BF (1984). Specification of the training and evaluation of cognitive therapists for outcome studies In Williams J & Spitzer R (Eds.), Psychotherapy research: Where are we and where should we go? (pp. 173–188). New York: Guilford Press.

Sholomskas DE, Syracuse-Siewert G, Rounsaville BJ, Ball SA, Nuro KF, & Carroll KM (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. [PubMed: 15709837]

Silverman WK, & Albano AM (1996). The anxiety disorders interview schedule for DSM-IV-child and parent versions. San Antonio, TX: Graywind Publications.

Simons AD, Padesky CA, Montemarano J, Lewis CC, Murakami J, Lamb K, ... & Beck AT (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(5), 751–756. [PubMed: 20873911]

Smith MM, McLeod BD, Southam-Gerow MA, Jensen-Doss A, Kendall PC, & Weisz JR (2016). Does the delivery of CBT for youth anxiety differ across research and practice settings? Behavior Therapy, 48(4), 501–516. [PubMed: 28577586]

Southam-Gerow MA, Chorpita BF, Miller LM, & Gleacher AA (2008). Are children with anxiety disorders privately referred to a university clinic like those referred from the public mental health system? Administration and Policy in Mental Health and Mental Health Services Research, 35, 168–180. [PubMed: 18046640]

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

Southam-Gerow MA, Weisz JR, Chu BC, McLeod BD, Gordis EB, & Connor-Smith JK (2010). Does cognitive behavioral therapy for youth anxiety outperform usual care in community clinics? An initial effectiveness test. Journal of the American Academy of Child and Adolescent Psychiatry, 49, 1043–52. [PubMed: 20855049]

Spilka MJ, & Dobson KS (2015). Promoting the internationalization of evidence-based practice: Benchmarking as a strategy to evaluate culturally transported psychological treatments. Clinical Psychology: Science and Practice, 22, 58–75.

Stern M, Niemann N, Wiedemann B, & Wenzlaff P (2011). Benchmarking improves quality in cystic fibrosis care: A pilot involving 12 centres. International Journal for Quality in Health Care, 23(3), 349–356. [PubMed: 21504960]

Webb CA, DeRubeis RJ, & Barber JP (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. Journal of Consulting and Clinical Psychology, 78(2), 200–211. [PubMed: 20350031]

Weersing VR (2005). Benchmarking the effectiveness of psychotherapy: Program evaluation as a component of evidence-based practice. Journal of the American Academy of Child & Adolescent Psychiatry, 44, 1058–1062. [PubMed: 16175111]

Weersing VR, & Weisz JR (2002). Community clinic treatment of depressed youth: Benchmarking usual care against CBT clinical trials. Journal of Consulting and Clinical Psychology, 70, 299–310. [PubMed: 11952188]

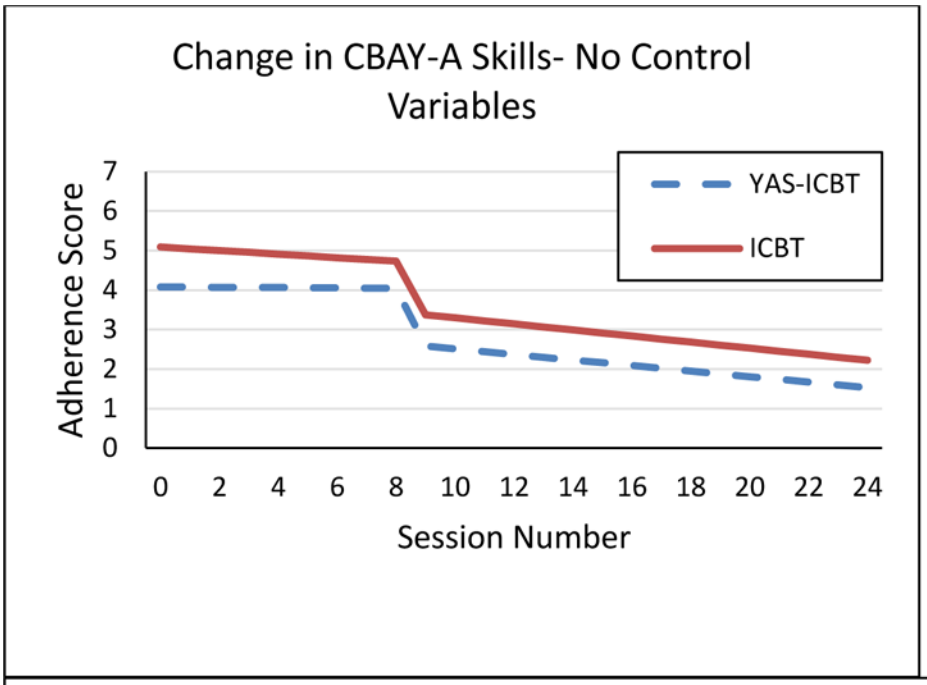
Weisz JR, Chorpita BF, Palinkas LA, Schoenwald SK, Miranda J, Bearman SK, Daleiden EL, Ugueto AM, Ho A, Martin J, Gray J, Alleyne A, Langer DA, Southam-Gerow MA, Gibbons RD, & the Research Network on Youth Mental Health. (2012). Testing standard and modular designs for psychotherapy with youth depression, anxiety, and conduct problems: A randomized effectiveness trial. Archives of General Psychiatry, 69, 274–282. [PubMed: 22065252]

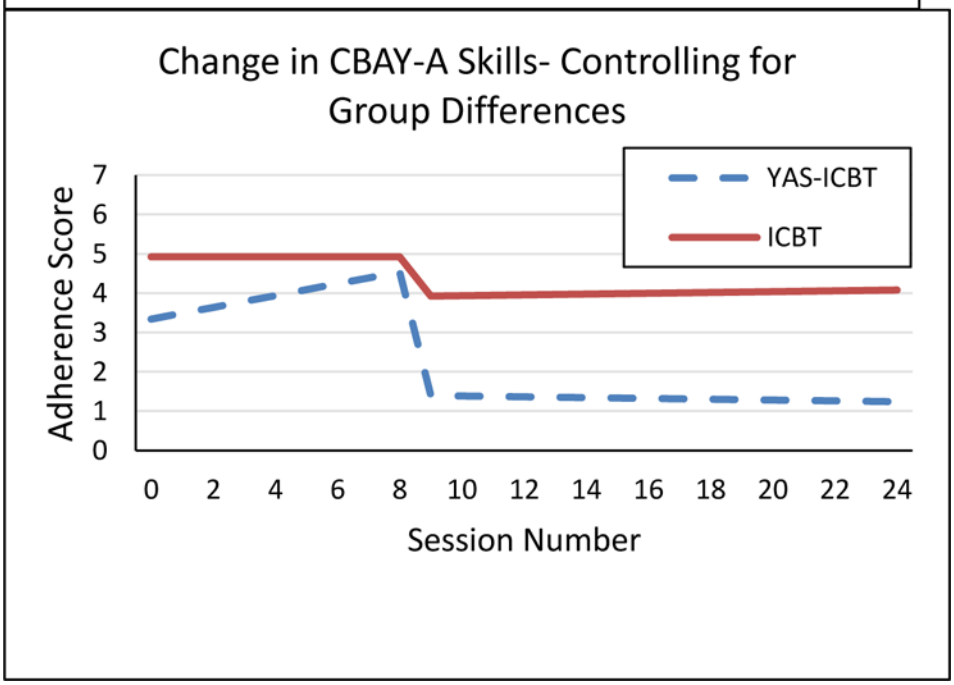
Weisz JR, & Kazdin AE (Eds.) (2010). Evidence-based psychotherapies for children and adolescents, 2nd *edition* New York: Guilford.nd

Weisz JR, Southam-Gerow MA, Gordis EB, Connor-Smith JK, Chu BC, Langer DA, . . . 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, 383–96. [PubMed: 19485581]

Wood JJ, Piacentini JC, Southam-Gerow M, Chu B, Sigman M (2006). Family cognitive behavioral therapy for child anxiety disorders. Journal of the American Academy of Child & Adolescent Psychiatry, 45, 314–321. [PubMed: 16540816]





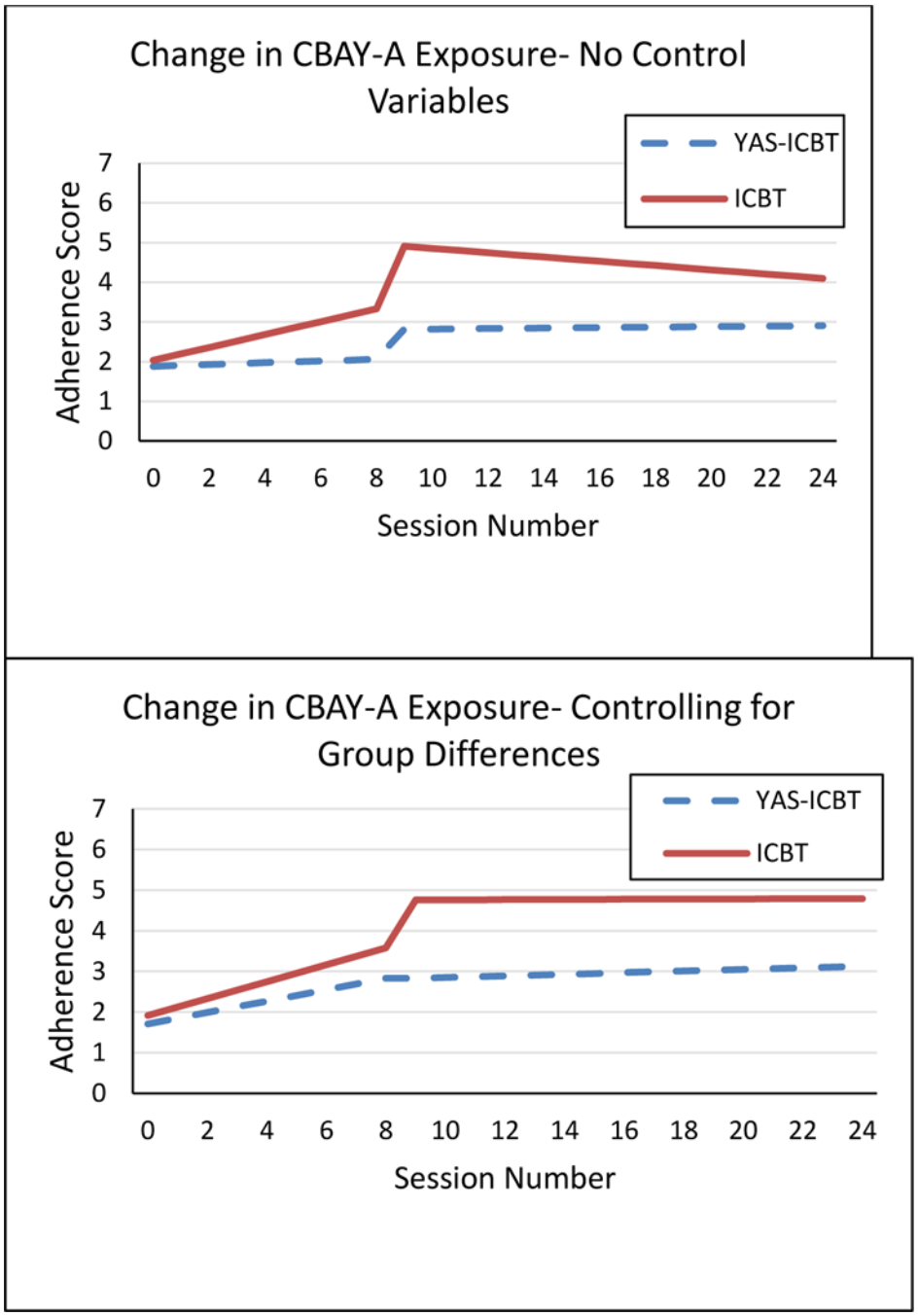


Figure 1.
Change over time in CBAY-A Skill and Exposure Scores

Table 1

Youth Descriptive Data and Comparisons Across Groups

Variable	*M (SD)* or %		*F* or Chi Square	*p*
	ICBT (N = 51)	YAS-ICBT (N = 17)		
Age	10.36 (1.90)	11.32 (2.32)	2.94	.091
Sex				
Male	60.78[a]	29.41	5.04	.025
Race/Ethnicity			15.48	.004
White	86.28[a]	41.18		
African-American	9.80	-		
Latino	1.96	17.65[b]		
Mixed/Other	1.96	5.88		
Not Reported	-	35.29[b]		
CBCL				
Total	63.18 (8.44)	64.19 (7.34)	0.183	.670
Internalizing	67.40 (8.37)	66.38 (8.33)	0.182	.671
Externalizing	52.96 (10.08)	60.81 (7.49)[b]	8.21	.006
Anxious-Depressed	62.94 (9.29)	68.63 (8.69)[b]	4.70	.034
Principal Anxiety Diagnoses			22.81	.0001
GAD	37.25[a]	5.88		
SAD	29.42	35.29		
SOP	33.33	23.54		
SP	-	35.29[b]		
Family Income				
Up to 60k per year	35.29	70.59[b]	7.92	.005
Number of Sessions per Case	15.92 (1.43)	16.82 (5.02)	1.36	.248
Weeks in Treatment per Case	19.52 (3.97)	26.38 (10.41)[b]	15.67	.0001
Length of Session in Minutes	52.93 (14.17)[a]	44.35 (11.05)	62.51	.0001
Total Minutes Spent in Treatment	842.93 (125.75)	750.14 (235.03)	4.33	.041
Number of Coded Sessions per Case	10.43 (2.84)	12.47 (4.61)[b]	4.71	.034

Note. ICBT = individual cognitive-behavioral therapy delivered in Kendall et al. study; YAS-ICBT = individual cognitive-behavioral therapy delivered in Youth Anxiety Study; 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 continuous variables.

[a] = ICBT > YAS-ICBT

[b] = YAS-ICBT > ICBT

Table 2

Therapist Descriptive Data and Comparisons Across Group

Variable	*M (SD)* or %		*F* or Chi Square	*p*
	ICBT (N = 16)	YAS-ICBT (N = 13)		
Sex				
Male	12.50	15.40	.104	.949
Race/Ethnicity			12.59	.05
White	81.25[a]	53.86		
African-American	-	-		
Asian-American	6.25	15.38		
Latino	6.25	15.38		
Mixed/Other	-	15.38		
Not Reported	6.25	-		
Professional Training			13.60	.001
Psychology	100.00[a]	38.48		
Social Worker	-	30.76[b]		
Other	-	30.76[b]		

Note. ICBT = individual cognitive-behavioral therapy delivered in Kendall et al. (2008) study; YAS-ICBT = ICBT delivered in YAS.

[a] = ICBT > YAS-ICBT

[b] = YAS-ICBT > ICBT

J Clin Child Adolesc Psychol. Author manuscript; available in PMC 2020 January 01.

Table 3

Multilevel Models of Adherence and Competence Between Groups

	Original Analyses				Analyses Controlling for Client Characteristics			
	Coefficient	**S.E.**	**p**	**ES**	**Coefficient**	**S.E.**	**p**	**ES**
CBAY-A Skills Subscale analyses								
Intercept (YAS-ICBT first skills session), γ_{000}	4.08	0.25	<.001		3.34	0.42	<.001	
Group Difference[1] in Intercept, γ_{010}	1.01	0.33	0.003	0.57	1.58	0.39	<.001	0.89
YAS-ICBT Linear Slope, γ_{100}	-0.004	0.027	0.883		0.15	0.064	0.020	
Group Difference in Linear Slope, γ_{110}	-0.04	0.036	0.27	-0.52	-0.15	0.057	0.012	-1.96
YAS-ICBT Change in Slope During Exposure, γ_{200}	-0.066	0.03	0.035		-0.16	0.080	0.050	
Group Difference in Slope Change, γ_{210}	0.034	0.044	0.44	0.15	0.17	0.077	0.029	0.77
YAS-ICBT Mean Difference Between Phases, γ_{300}	-0.87	0.35	0.02		-1.86	0.60	0.004	
Group Difference in Mean Between Phases, γ_{310}	-0.15	0.42	0.73	-0.08	.78	0.60	0.200	0.44
CBAY-A Exposure Subscale analyses								
Intercept (YAS-ICBT first exposure session), γ_{000}	2.06	.29	<.001		2.83	0.54	<.001	
Group Difference in Intercept, γ_{010}	1.27	.36	<.001	1.17	0.75	0.51	0.15	0.69
YAS-ICBT Linear Slope, γ_{100}	.022	.029	.447		0.14	0.069	0.058	
Group Difference in Linear Slope, γ_{110}	.14	.038	<.001	2.95	0.068	0.058	0.25	1.43
YAS-ICBT Change in Slope During Exposure, γ_{200}	-.017	.051	.752		-0.12	0.097	0.24	
Group Difference in Slope Change, γ_{210}	-.20	.067	.005	-1.47	-0.086	0.090	0.35	-0.63
YAS-ICBT Mean Difference Between Phases, γ_{300}	.75	.34	.038		-.019	0.59	0.97	
Group Difference in Mean Between Phases, γ_{310}	.88	.43	.043	0.81	1.20	0.62	0.061	1.10
CBAY-C Total scale analyses								
Intercept (YAS-ICBT mean across sessions), γ_{000}	3.71	.19	<.001		3.81	0.24	<.001	
Group Difference in Intercept, γ_{010}	1.53	.25	<.001	1.13	1.48	0.28	<.001	1.09

Note.ICBT = individual cognitive-behavioral therapy delivered in Kendall et al. (2008) study; YAS-ICBT = ICBT delivered in YAS; S.E. = standard error; ES = effect size. Group differences were examined with a dummy coded variable coded as ICBT (1) vs YAS-ICBT (0)