



## Examining How CBT Interventions for Anxious Youth Are Delivered Across Settings

Julia R. Cox

Bryce D. McLeod

Virginia Commonwealth University

Amanda Jensen-Doss

University of Miami

Vishnupriya Srivastava

Michael A. Southam-Gerow

Virginia Commonwealth University

Philip C. Kendall

Temple University

John R. Weisz

Harvard University

Most efforts to assess treatment integrity—the degree to which a treatment is delivered as intended—have conflated content (i.e., therapeutic interventions) and delivery (i.e., strategies for conveying the content, such as modeling). However, there may be value in measuring content and delivery separately. This study examined whether the quantity (how much) and quality (how well) of delivery strategies for individual cognitive behavioral therapy (ICBT) for youth anxiety varied when the same evidence-based treatment was implemented in research and community settings. Therapists ( $N = 29$ ; 69.0% White; 13.8% male) provided ICBT to 68 youths ( $M$  age = 10.60 years,

$SD = 2.03$ ; 82.4% white; 52.9% male) diagnosed with a principal anxiety disorder in research or community settings. Training and supervision protocols for therapists were comparable across settings. Two independent teams of trained coders rated 744 sessions using observational instruments designed to assess the quantity and quality of delivery of interventions found in ICBT approaches. Overall, both the quantity and quality of delivery of interventions found in ICBT approaches were significantly lower in the community settings. The extent to which didactic teaching, collaborative teaching, and rehearsal were used systematically varied over the course of treatment. In general, differences in the quantity and quality of delivery observed between settings held when differences in youth characteristics between settings were included in the model. Our findings suggest the potential relevance of measuring how therapists deliver treatment separate from the content.

This work was supported by the National Institute of Mental Health (R01 MH086529).

Address correspondence to Julia R. Cox, Department of Psychology, Virginia Commonwealth University, 806 West Franklin Street, Box 842018, Richmond, VA 23284; e-mail: [coxjr4@vcu.edu](mailto:coxjr4@vcu.edu).

0005-7894/© 2019 Association for Behavioral and Cognitive Therapies. Published by Elsevier Ltd. All rights reserved.

**Keywords:** cognitive-behavioral treatment; youth anxiety; treatment delivery; treatment integrity

EFFORTS TO IMPROVE THE quality of mental health services have focused, in part, on implementing evidence-based treatments (EBTs) in community mental health settings. EBTs, typically developed and evaluated in university-based settings (called *research settings*), can be challenging to implement in community mental health settings (called *community settings*). For example, differences have been reported between the demographic and clinical characteristics of youth accessing treatment in research and community settings (Ehrenreich-May et al., 2011; Southam-Gerow, Chorpita, Miller, & Gleacher, 2008). Findings indicate that there are differences in the conditions within which treatment is carried out in research versus community settings (Weisz, Krumholz, Santucci, Thomassin, & Ng, 2015). Meta-analyses have shown that EBTs, which are typically developed and initially tested in research settings, on average show a decline in effect size when later tested against usual clinical care in more clinically representative and community settings (Weisz et al., 2013). Taken together, these findings highlight a need to understand how implementation of EBTs may differ in research versus community settings. One dimension along which research and community contexts may be compared is treatment delivery—defined here as strategies for conveying therapeutic concepts and content (McLeod et al., 2019; Smith et al., 2017). Identifying differences in how an EBT is delivered across settings and over time has potential to help us identify useful EBT adaptations or quality improvement targets (Garland et al., 2010; Lyon & Koerner, 2016).

Most efforts to measure therapist behavior have conflated content and delivery (Morgan, Davis, Richardson, & Perkins, 2018). Content focuses on the therapeutic interventions found in a treatment protocol deemed critical for symptom reduction that convey new knowledge, skills, and experiences (McLeod et al., 2018; Morgan et al., 2018), such as cognitive restructuring or exposure tasks. In contrast, delivery (or *process*; Morgan et al., 2018) focuses on *how* the content is delivered to a client (Garland et al., 2010). For example, a therapist may choose to present information about cognitive restructuring in a collaborative manner by asking a youth open-ended questions or model cognitive restructuring by providing examples. How content is delivered may influence client learning and skill acquisition (Morgan et al., 2018); thus, it may be important for instruments designed to assess therapist behavior to separate delivery from content (Garland et al., 2010; Southam-Gerow et al., 2016).

To accomplish the separation of delivery from content in cognitive-behavioral treatments, one can conceptualize the delivery of content along a contin-

uum from passive to active (e.g., Morgan et al., 2018). Passive delivery involves therapist-driven presentation of new information and skills, asks for little client involvement, and includes didactic teaching strategies (i.e., therapist provides the client with new information while the client receives the information in a passive manner) and modeling (i.e., therapist demonstrates to the client how to perform a specific skill while the client observes in a passive manner). In contrast, active delivery involves encouraging the client to engage with material and includes collaborative teaching (i.e., teaching new information using open-ended questions and guided discovery, requiring client involvement) and rehearsal (i.e., client practices a skill).

Theoretical and empirical work with adults and children suggests that passive and active delivery strategies may serve different functions (e.g., Bornstein, Bellack, & Hersen, 1977; Ross, 1964). Passive delivery strategies are often employed to provide information that is new or corrective, such as introducing a new cognitive-behavioral skill; within learning theory, one can draw parallels to the information assimilation process described by Coleman (1977). In contrast, active delivery strategies may be important for learning how to apply new knowledge and skills, such as practicing newly acquired cognitive-behavioral skills; within learning theory, these strategies resemble experiential learning (Coleman, 1977; Kolb, Boyatzis, & Mainemelis, 2000). Although treatment content influences what skills and knowledge are taught, delivery may contribute to how well the new knowledge and skills are acquired by the client over treatment.

Understanding if the delivery of content differs across contexts may help us determine which facets of EBTs are different when implemented across settings. Studies that measure treatment integrity have shown that treatment adherence (i.e., extent to which treatment protocol interventions were delivered; Perepletchikova & Kazdin, 2005) and competence (i.e., level of responsiveness and skill with which a therapist delivers the treatment protocol interventions; Barber, Sharpless, Klostermann, & McCarthy, 2007) differ across settings (e.g., McLeod et al., 2019; Smith et al., 2017). Differences may also exist for the delivery of that content. It is, for example, plausible that differences in delivery exist across settings because some therapists are less comfortable delivering content that requires active delivery, such as exposure for anxiety (e.g., Becker, Zayfert, & Anderson, 2004; Deacon et al., 2013). This study investigated whether differences exist in the quantity and quality of delivery when the same EBT is delivered in research and community settings. Our investigation

focused on Coping Cat, an individual cognitive-behavioral therapy (ICBT) for youth anxiety (Kendall & Hedtke, 2006). Coping Cat is ideal for this study because it is an efficacious program that has been tested in research and community settings. We examined treatment process data from two randomized controlled trials (RCTs) that employed Coping Cat: one in a research setting (Kendall, Hudson, Gosch, Flannery-Schroeder, & Suveg, 2008) and one in community clinic settings (Southam-Gerow et al., 2010). We hypothesized that adherence scores for active delivery of ICBT interventions (i.e., rehearsal, collaborative teaching) would be lower in the community than research setting. We also hypothesized that competence scores for active ICBT delivery would be lower in community than research setting. Finally, we explored whether the quantity and quality of delivery changed over the course of treatment and whether pattern of change over treatment differed by setting.

## Method

### DATA SOURCES AND PARTICIPANTS

The data for the present study were collected from 68 youth who participated in one of two RCTs. The first efficacy RCT compared ICBT, family-CBT, and an active control group (Kendall et al., 2008). The second effectiveness RCT compared ICBT (YAS-ICBT) to usual care in the Youth Anxiety Study (YAS; Southam-Gerow et al., 2010). This study only included the ICBT groups (ICBT, YAS-ICBT). Treatment data were recorded sessions collected in each RCT. Inclusion criteria for this study required youth to have: (a) at least two recorded sessions, and (b) received ICBT from a single therapist (see Kendall et al., 2008; Southam-Gerow et al., 2010).

### Research Setting

In Kendall et al. (2008), 51 youth ( $M$  age = 10.36 years,  $SD$  = 1.90; 86.3% White; 60.8% male) received ICBT at a research clinic that specialized in the treatment of anxiety disorders in the city of Philadelphia. Therapists ( $n$  = 16; 12.5% male) were 81.3% White, 6.3% Latino, 6.3% Asian/Pacific Islander, and 6.3% did not report their ethnicity/racial background. Therapists were either licensed clinical psychologists or clinical psychology doctoral trainees. Posttreatment, 64.0% of youth no longer met diagnostic criteria for their principal anxiety disorder (see Kendall et al., 2008).

### Community Settings

In Southam-Gerow et al. (2010), 17 youth ( $M$  age = 11.32 years,  $SD$  = 2.32; 41.2% White; 29.4% male)

received treatment at community-based mental health clinics across Los Angeles County. Therapists were employed by the clinics, volunteered to participate, and were randomly assigned to treatment groups. The 13 YAS-ICBT therapists (15.4% male) were 53.9% White, 15.4% Latino, 15.4% Asian/Pacific Islander, and 15.4% mixed/other. Professional composition was 30.8% social workers, 46.3% psychology (30.8% master's level, 15.4% doctoral level), and 23.1% reported "other" degree. Posttreatment, 66.7% of youths in the YAS-ICBT condition no longer met diagnostic criteria (see Southam-Gerow et al., 2010).

### INDIVIDUAL COGNITIVE BEHAVIORAL THERAPY

Therapists in both studies delivered the Coping Cat program, an ICBT program for youth with anxiety disorders (Kendall & Hedtke, 2006). Coping Cat consists of 16 sessions of which 14 sessions are conducted with the youth and two sessions are held with the caregivers. Coping Cat is divided into two phases that require both passive and active strategies that require both passive and active strategies to deliver interventions. The first eight sessions (i.e., skill-building phase) focus on anxiety management skills training and includes psychoeducation, emotion education, building a fear ladder, relaxation, cognitive restructuring, problem solving, self-reward, and building a coping plan. The latter eight sessions (i.e., exposure phase) focus on gradual exposure. Homework is assigned to the youth throughout the program. In both studies, quality control procedures included the same three components, including a treatment protocol, a training workshop, and weekly supervision (Sholomskas et al., 2005). The therapist training and supervision in both RCTs was provided by experts in the Coping Cat program (i.e., treatment developer or postdoctoral trainee who was trained by the treatment developer). Adherence to Coping Cat was measured with the Coping Cat Brief Adherence Scale in both RCTs (see Kendall, 1994; Kendall et al., 1997). Based on this scale, the therapists in both studies demonstrated more than 90.0% general adherence to the Coping Cat protocol (see Kendall et al., 2008; Southam-Gerow et al., 2010, for details).

### ADHERENCE AND COMPETENCE INSTRUMENTS

*CBT Adherence Scale for Youth Anxiety (CBAY-A; Southam-Gerow et al., 2016)*

The CBAY-A is a 22-item observer-rated instrument that assesses three domains: (a) *Standard*, 4

items that represent standard CBT interventions (e.g., Homework assigned), (b) *Model*, 12 items that assess content commonly found in ICBT programs for youth anxiety (e.g., Psychoeducation, Cognitive, Problem Solving, Self-Reward, Exposure; see Southam-Gerow et al., 2016, for additional details), and (c) *Delivery*, 6 items that measure how model items were delivered (i.e., Didactic Teaching, Collaborative Teaching, Modeling, Rehearsal, Self-Disclosure, Coaching). After watching the entirety of a treatment session, coders rate each item on a 7-point extensiveness scale that represents frequency and thoroughness; anchors are: 1 = *not at all*, 3 = *somewhat*, 5 = *considerably*, 7 = *extensively*. The Model item scores have shown evidence of construct validity; they discriminated between therapists delivering ICBT across research and community settings from therapists delivering usual care (McLeod et al., 2018; Southam-Gerow et al., 2016). The current study uses four CBAY-A Delivery items: Didactic Teaching, Collaborative Teaching, Modeling, and Rehearsal (Coaching and Self-Disclosure items were dropped due to poor score validity; see McLeod et al., 2018). For the analyses, Didactic Teaching and Modeling were considered passive delivery strategies, and Collaborative Teaching and Rehearsal were considered active delivery strategies. For this sample, the average item inter-rater reliability of the four items, ICC(2,2), was 0.73 ( $SD = 0.08$ ) and ranged from 0.68 to 0.84.

#### *CBT for Youth Anxiety Competence Scale (CBAY-C; McLeod et al., 2018)*

The CBAY-C is a 25-item observer-rated instrument. The CBAY-C is similar in format and content to the CBAY-A; it assesses the same three domains: (a) *Standard* (4 items), (b) *Model* (12 items), (c) *Delivery* (6 items), and also includes (d) *Global*, 2 items that assess level of skillfulness and responsiveness of therapist CBT delivery overall. After watching the entirety of a treatment session, coders make ratings on a 7-point competence scale that represent skillfulness and responsiveness; anchors are: 1 = *very poor*; 3 = *acceptable*; 5 = *good*; 7 = *excellent*. The CBAY-C Model items have shown evidence of construct validity (McLeod et al., 2018; McLeod et al., 2019). The current study uses the four CBAY-C Delivery items that correspond with included CBAY-A items. Correlations between the four CBAY-C and CBAY-A Delivery items ranged from .31 to .56 (see McLeod et al., 2018). For the present sample, the average item inter-rater reliability, ICC(2,2), was .63 ( $SD = 0.08$ ) representing “good” inter-rater agreement (Cicchetti, 1994) and ranged from .49 to .69.

#### ASSESSMENTS COLLECTED IN THE ORIGINAL RCTS

Diagnostic and symptom tools collected in the original RCTs were used in this study for control and setting comparison purposes. Kendall et al. (2008) used the *Anxiety Disorders Interview Schedule for DSM-IV: Child and Parent Versions* (Silverman & Albano, 1996) to evaluate youth DSM-IV disorders. Southam-Gerow et al. (2010) used the *Diagnostic Interview Schedule for Children Version 4.0* (Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000) to evaluate youth DSM-IV disorders. The *Child Behavior Checklist* (CBCL; Achenbach, 1991) was used in both studies to evaluate youth symptoms on broad-band scales (e.g., Externalizing) and narrow-band (e.g., Somatic Complaints) subscales. In this study,  $t$ -scores from the Total scale and three subscales were used: (1) Internalizing (broad-band), (2) Externalizing (broad-band), and (3) Anxious-Depressed (narrow-band).

#### CODING AND SESSION SAMPLING PROCEDURES

Coders were four clinical psychology doctoral students (25.0% male,  $M$  age = 28.00 years,  $SD = 2.71$ ; 50.0% Latina, 50.0% white); two coded the CBAY-A and two coded the CBAY-C. Coders were trained over 3 months, followed by weekly meetings to prevent drift. Coders were naïve to hypotheses and sessions were assigned randomly to each coder. Each session was double coded; we used mean scores in analyses to reduce measurement error. For each treatment case, all available sessions were coded, except the first and last session, as these sessions were likely to comprise intake or termination content (e.g., assessment). Of the 1,098 sessions held, 744 (67.7%) were rated (65.5% ICBT [ $n = 532$ ], 74.1% YAS-ICBT [ $n = 212$ ]).

#### DATA ANALYSIS PLAN

Multilevel modeling (Raudenbush & Bryk, 2002) was used for analyses of setting differences in the quantity and quality scores and were conducted using HLM 7.01 (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011) to account for nesting (i.e., youth in therapists, repeated measures in youth). A model fitting process compared three ways of modeling each dependent variable: (1) an intercept-only model (i.e., no change over time), (2) a linear model, and (3) a piecewise linear model using an increment/decrement model (Raudenbush & Bryk, 2002) that included a base rate slope parameter estimating change over all sessions, an

increment/decrement parameter that estimated change in slope after the exposure phase started, and a parameter that examined mean differences in slopes between the skills and exposure phases (described earlier in our description of ICBT). The piecewise linear model was designed to ascertain if treatment delivery varied across the skill-building and the exposure phases of Coping Cat, since previous research has indicated that treatment adherence varies across them (McLeod et al., 2019). Change in deviance statistics, AIC, and BIC values were examined to determine the best fitting model for each dependent variable. Below is an example of the piecewise model that was fit to scores on the CBAY-A Didactic Teaching item, 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 YAS-ICBT is the setting difference variable (YAS-ICBT = 1, IBCT = 0):

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

$$L2: \pi_{0jk} = \beta_{00k} + \beta_{01k}^*(YAS-ICBT_{jk}) + r_{0jk}$$

$$\pi_{1jk} = \beta_{10k} + \beta_{11k}^*(YAS-ICBT_{jk}) + r_{1jk}$$

$$\pi_{2jk} = \beta_{20k} + \beta_{21k}^*(YAS-ICBT_{jk}) + r_{2jk}$$

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

$$L3: \beta_{00k} = \gamma_{000} + u_{00k}$$

$$\beta_{01k} = \gamma_{010}$$

$$\beta_{10k} = \gamma_{100} + u_{10k}$$

$$\beta_{11k} = \gamma_{110}$$

$$\beta_{20k} = \gamma_{200} + u_{20k}$$

$$\beta_{21k} = \gamma_{210}$$

$$\beta_{30k} = \gamma_{300} + u_{30k}$$

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

For the linear model, the model above was modified to only include TIMEWEEK as a predictor, and the intercept only model did not include any level 1 predictors.

To estimate the magnitude of the between-group differences, we calculated Cohen’s *d* effect sizes. In order to calculate effect sizes for parameters representing mean differences the setting difference parameter estimates were divided by the raw data standard deviations. Following Feingold (2009), to calculate effect sizes for the slope parameter, we first multiplied the slope parameter by treatment length (or, in the case of the increment/decrement slope, by the length of exposure) and then divided the product by the raw standard deviation. Because treatment length differed across settings (see Table 1), the overall average length of treatment (22.56 weeks [*SD* = 9.80]) and the overall average length of the exposure phase (9.47 weeks [*SD* = 4.25]) were used. Effect sizes

Table 1  
Youth Descriptive Data and Comparisons Across Settings

| Variable                 | <i>M (SD) or %</i>    |                            | <i>F or Chi Square</i> | <i>P</i> |
|--------------------------|-----------------------|----------------------------|------------------------|----------|
|                          | ICBT ( <i>N</i> = 51) | YAS-ICBT ( <i>N</i> = 17)  |                        |          |
| Age                      | 10.36 (1.90)          | 11.32 (2.32)               | 2.94                   | .091     |
| Sex                      |                       |                            |                        |          |
| Male                     | 60.8 <sup>a</sup>     | 29.4                       | 5.04                   | .025     |
| Race/Ethnicity           |                       |                            | 15.48                  | .004     |
| White                    | 86.3 <sup>a</sup>     | 41.2                       |                        |          |
| African-American         | 9.80                  | -                          |                        |          |
| Latino                   | 1.9                   | 17.6 <sup>b</sup>          |                        |          |
| Mixed/Other              | 1.9                   | 5.9                        |                        |          |
| Not Reported             | -                     | 35.3 <sup>b</sup>          |                        |          |
| CBCL                     |                       |                            |                        |          |
| Total                    | 62.92 (8.56)          | 64.19 (7.34)               | 0.28                   | .596     |
| Internalizing            | 67.08 (8.60)          | 66.38 (8.33)               | 0.08                   | .775     |
| Externalizing            | 52.76 (10.08)         | 60.81 (7.49) <sup>b</sup>  | 8.65                   | .005     |
| Anxious-Depressed        | 62.94 (9.29)          | 68.63 (8.69) <sup>b</sup>  | 4.70                   | .034     |
| Principal Diagnoses      |                       |                            | 22.81                  | .0001    |
| GAD                      | 37.3 <sup>a</sup>     | 5.9                        |                        |          |
| SAD                      | 29.4                  | 35.3                       |                        |          |
| SOP                      | 33.3                  | 23.5                       |                        |          |
| SP                       | -                     | 35.3 <sup>b</sup>          |                        |          |
| Family Income            |                       |                            |                        |          |
| Up to 60k per year       | 35.3                  | 70.6 <sup>b</sup>          | 7.92                   | .005     |
| Number of Sessions       | 15.92 (1.43)          | 16.82 (5.02)               | 1.36                   | .248     |
| Weeks in Treatment       | 19.52 (3.97)          | 26.38 (10.41) <sup>b</sup> | 15.67                  | .0001    |
| Number of Coded Sessions | 10.43 (2.84)          | 12.47 (4.61) <sup>b</sup>  | 4.71                   | .034     |

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

<sup>a</sup> = ICBT > YAS-ICBT  
<sup>b</sup> = YAS-ICBT > ICBT

reported here are Cohen’s *d* (1988) and we use established guidelines to guide interpretation, such that .20 is a small effect, .50 is a medium effect, and .80 is a large effect. We also examined whether the setting differences remained significant when controlling for youth characteristics that differed between settings. Control variables were entered simultaneously into each model, and we grand mean centered continuous variables.

## Results

### PRELIMINARY ANALYSES

Our first step was to conduct sample bias analyses to determine if the 68 youth and 29 therapist participants selected for this study differed from the other participants in the parent studies (Kendall et al., 2008; Southam-Gerow et al., 2010). We found a lower proportion of Black youth (0% vs. 16.7%) and a higher proportion of White youth (41.2% vs. 29.2%) in the YAS-ICBT sample compared to the parent study,  $\chi^2(3, N = 24) = 11.53, p = .009$ .

Missing data patterns were examined. Of the 1,098 sessions held, 744 (67.7%) were rated; on average, 65.5% ( $SD = 17.4\%$ ) of each ICBT case was coded ( $n = 532$  of 812 sessions) and 74.3% ( $SD = 15.8\%$ ) of each YAS-ICBT case was coded ( $n = 212$  of 286 sessions). There was no significant difference between settings in terms of the percent of sessions coded,  $t(66) = 1.85, p = .07$ , or in the percent of sessions coded from the first and second half of treatment (first half = 67.6%; second half 67.9%;  $t(67) = 0.07, p = .95$ ). For the youth-level control variables, rates of missing data were 8.8% (youth ethnicity, family income data) or less. These data were missing completely at random (Little's MCAR test  $X^2 = 236.8, DF = 237, p = .49$ ). Analyses involving the youth-level variables described below were conducted using the multiple imputation function in HLM 7.01. Ten datasets were imputed using SPSS Version 25.0 (IBM Corp, 2017).

We examined differences between settings (ICBT, YAS-ICBT) on several youth demographic, baseline clinical, and treatment characteristics (see Table 1). Settings differed on sex, race/ethnicity, level of externalizing symptomatology, level of anxiety and depressive symptomatology (as measured by the CBCL Anxious-Depressed subscale), principal anxiety disorders, family income level, and weeks in treatment, so these variables were selected as control variables. As in Table 2, settings did not differ on therapist sex or race/ethnicity but did differ on therapist professional training. Because therapist professional training was confounded with setting, we did not evaluate this as a control variable.

### SETTING COMPARISONS: ADHERENCE AND COMPETENCE

#### Didactic Teaching

For the CBAY-A Didactic Teaching item, the piecewise model was the best fitting model. In the base CBAY-A Didactic Teaching piecewise model (i.e., without setting difference predictors), most of the variability in the model was at the therapist level

Table 2  
Therapist Descriptive Data and Comparisons Across Setting

| Variable              | <i>M (SD) or %</i>       |                              | <i>F or Chi Square</i> | <i>p</i> |
|-----------------------|--------------------------|------------------------------|------------------------|----------|
|                       | ICBT<br>( <i>N</i> = 16) | YAS-ICBT<br>( <i>N</i> = 13) |                        |          |
| Sex                   |                          |                              |                        |          |
| Male                  | 12.50                    | 15.40                        | .05                    | .823     |
| Race/Ethnicity        |                          |                              | 5.12                   | .266     |
| White                 | 81.25 <sup>a</sup>       | 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 |                          |                              | 11.36                  | .003     |
| Psychology            | 100.00 <sup>a</sup>      | 46.25                        |                        |          |
| Social Worker         | -                        | 30.77 <sup>b</sup>           |                        |          |
| Other                 | -                        | 23.08 <sup>b</sup>           |                        |          |

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

<sup>a</sup> = ICBT > YAS-ICBT.

<sup>b</sup> = YAS-ICBT > ICBT.

(64.7% of the variability in intercept, 93.7% of the variability in linear slope, 92.3% of the variability in slope change, and 74.6% of the variability in the mean difference between phases). Model results examining setting differences are presented in Table 3. At the first session, scores on the CBAY-A Didactic Teaching item were lower in the YAS-ICBT setting than ICBT ( $\gamma_{010} = -.88, p = .001, d = -.73$ ). There was no significant linear change over time or a difference in the linear slope after the beginning of exposure, and no setting differences on these two model parameters. However, scores for both settings were lower for the CBAY-A Didactic Teaching item during the exposure phase compared to the skills phase, although that decrease was less pronounced in the YAS-ICBT setting ( $\gamma_{300} = .73, p = .011, d = .60$ ). See Fig. 1.

For scores on the CBAY-C Didactic Teaching item, the intercept only model was the best fitting model, indicating no systematic change over time. Nearly all variability in the CBAY-C Didactic Teaching item scores was at the therapist (44.59%) or session (55.39%) level, with basically no variance at the youth level (0.02%). The YAS-ICBT setting had significantly lower CBAY-C Didactic Teaching scores than ICBT ( $\gamma_{010} = -1.33, p < .001, d = -1.17$ ).

#### Collaborative Teaching

For scores on the CBAY-A and CBAY-C Collaborative Teaching items, the intercept only model was the best fitting model, indicating no systematic

Table 3  
Multilevel Models of Adherence and Competence Between Settings

|   | Original Analyses |       |       |       | Analyses Controlling for Client Characteristics |       |       |       |
|---|-------------------|-------|-------|-------|---|-------|-------|-------|
|   | Coefficient       | S.E.  | p     | ES    | Coefficient                                     | S.E.  | p     | ES    |
| <b>CBAY-A Didactic Teaching</b>   |                   |       |       |       |   |       |       |       |
| Intercept (ICBT first session), $\gamma_{000}$                          | 3.25              | 0.15  | <.001 |       | 3.1   | 0.16  | <.001 |       |
| Setting Difference <sup>1</sup> in Intercept, $\gamma_{010}$            | -0.88             | 0.25  | 0.001 | -0.73 | -0.44   | 0.39  | 0.27  | 0.36  |
| ICBT Linear Slope, $\gamma_{100}$                                       | -0.00054          | 0.017 | 0.98  |       | -0.03   | 0.019 | 0.12  |       |
| Setting Difference in Linear Slope, $\gamma_{110}$                      | -0.0078           | 0.023 | 0.737 | 0.15  | 0.075   | 0.045 | 0.11  | -1.43 |
| ICBT Change in Slope During Exposure, $\gamma_{200}$                    | -0.017            | 0.025 | 0.49  |       | 0.0026  | 0.03  | 0.93  |       |
| Setting Difference in Slope Change, $\gamma_{210}$                      | -0.016            | 0.033 | 0.62  | -0.28 | -0.082  | 0.065 | 0.22  | 0.64  |
| ICBT Mean Difference Between Skills and Exposure Phases, $\gamma_{300}$ | -1.33             | 0.16  | <.001 |       | -1.18   | 0.17  | <.001 |       |
| Setting Difference in Mean Difference Between Phases, $\gamma_{310}$    | 0.73              | 0.27  | 0.00  | 0.60  | 0.21  | 0.42  | 0.62  | -0.17 |
| <b>CBAY-C Didactic Teaching</b>   |                   |       |       |       |   |       |       |       |
| Intercept (ICBT mean across sessions), $\gamma_{000}$                   | 4.78              | 0.12  | <.001 |       | 4.78  | 0.14  | <.001 |       |
| Setting Difference in Intercept, $\gamma_{010}$                         | -1.33             | 0.21  | <.001 | -1.17 | -1.38   | 0.28  | <.001 | 1.21  |
| <b>CBAY-A Collaborative Teaching</b>                                    |                   |       |       |       |   |       |       |       |
| Intercept (ICBT mean), $\gamma_{000}$                                   | 3.69              | 0.14  | <.001 |       | 3.73  | 0.14  | <.001 |       |
| Setting Difference <sup>1</sup> in Intercept, $\gamma_{010}$            | -0.73             | 0.22  | 0.002 | -0.48 | -0.92   | 0.29  | 0.004 | 0.60  |
| <b>CBAY-C Collaborative Teaching</b>                                    |                   |       |       |       |   |       |       |       |
| Intercept (ICBT mean), $\gamma_{000}$                                   | 5.19              | 0.14  | <.001 |       | 5.2   | 0.15  | <.001 |       |
| Setting Difference in Intercept, $\gamma_{010}$                         | -1.46             | 0.23  | <.001 | -1.19 | -1.55   | 0.3   | <.001 | 1.26  |
| <b>CBAY-A Modeling</b>  |                   |       |       |       |   |       |       |       |
| Intercept (ICBT first session), $\gamma_{000}$                          | 2.19              | 0.17  | <.001 |       | 2.2   | 0.19  | <.001 |       |
| Setting Difference <sup>1</sup> in Intercept, $\gamma_{010}$            | -0.32             | 0.27  | 0.25  | -0.25 | -0.38   | 0.39  | 0.32  | -0.29 |
| ICBT Linear Slope, $\gamma_{100}$                                       | -0.069            | 0.02  | 0.00  |       | -0.071  | 0.021 | 0.003 |       |
| Setting Difference in Linear Slope, $\gamma_{110}$                      | 0.028             | 0.031 | 0.375 | 0.86  | -0.017  | 0.054 | 0.76  | -0.30 |
| ICBT Change in Slope During Exposure, $\gamma_{200}$                    | 0.057             | 0.03  | 0.06  |       | 0.019   | 0.035 | 0.18  |       |
| Setting Difference in Slope Change, $\gamma_{210}$                      | -0.039            | 0.043 | 0.37  | -0.29 | 0.034   | 0.075 | 0.65  | 0.25  |
| ICBT Mean Difference Between Skills and Exposure Phases, $\gamma_{300}$ | -0.48             | 0.15  | 0.003 |       | -0.47   | 0.17  | 0.012 |       |
| Setting Difference in Mean Difference Between Phases, $\gamma_{310}$    | 0.15              | 0.26  | 0.57  | 0.12  | -0.01   | 0.45  | 0.83  | 0.01  |
| <b>CBAY-C Modeling</b>  |                   |       |       |       |   |       |       |       |
| Intercept (ICBT mean), $\gamma_{000}$                                   | 5.12              | 0.12  | <.001 |       | 5.09  | 0.12  | <.001 |       |
| Setting Difference in Intercept, $\gamma_{010}$                         | -0.87             | 0.21  | <.001 | -1.05 | -0.73   | 0.3   | 0.02  | -0.88 |
| <b>CBAY-A Rehearsal</b>   |                   |       |       |       |   |       |       |       |
| Intercept (ICBT first session), $\gamma_{000}$                          | 3.72              | 0.16  | <.001 |       | 3.61  | .18   | <.001 |       |
| Setting Difference <sup>1</sup> in Intercept, $\gamma_{010}$            | -0.89             | 0.26  | 0.001 | -0.46 | -0.79   | 0.40  | 0.064 | -0.41 |
| ICBT Linear Slope, $\gamma_{100}$                                       | 0.099             | 0.015 | <.001 |       | 0.10  | 0.017 | <.001 |       |
| Setting Difference in Linear Slope, $\gamma_{110}$                      | -0.11             | 0.026 | <.001 | -1.32 | -0.064  | 0.043 | 0.152 | -0.77 |
| <b>CBAY-C Rehearsal</b>   |                   |       |       |       |   |       |       |       |
| Intercept (ICBT mean), $\gamma_{000}$                                   | 5.27              | 0.12  | <.001 |       | 3.85  | 0.021 | <.001 |       |
| Setting Difference in Intercept, $\gamma_{010}$                         | -1.41             | 0.2   | <.001 | -1.31 | -1.42   | 0.27  | <.001 | -1.31 |

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. Setting differences were examined with a dummy coded variable coded as YAS-ICBT (1) vs ICBT (0)

change over time. Nearly all variability in scores was either at the therapist (CBAY-A: 15.3%; CBAY-C: 48.8%) or session (CBAY-A: 83.4%; CBAY-C: 51.1%) level, with little to no variance at the youth level (CBAY-A: 1.3%; CBAY-C: 0.07%). There were significant setting differences, with the YAS-ICBT setting showing lower Collaborative

Teaching CBAY-A ( $\gamma_{010} = -.73, p = .002, d = -.48$ ) and CBAY-C ( $\gamma_{010} = -1.46, p < .001, d = -1.19$ ) scores. See Fig. 2.

#### Modeling

The piecewise model was the best fitting model for scores on the CBAY-A Modeling item, with more

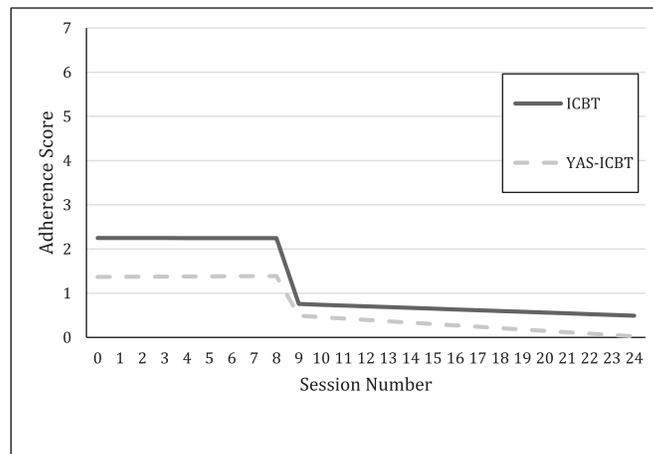


FIG. 1 Change in CBAY-A Didactic Teaching.

variability at the therapist level (68.9% of the variability in intercept, 48.4% of the variability in linear slope, 64.9% of the variability in slope change, and 53.8% of the variability in the mean difference between phases). As depicted in Fig. 3, there was a significant linear decrease over time in scores on Modeling in the ICBT setting, as well as a drop in mean levels of scores on Modeling after the beginning of the exposure phase; there were no significant setting differences (Table 3). For scores on the CBAY-C Modeling item, the intercept only model was the best fitting model, and nearly all of the variability was at the therapist (43.5%) or session (56.5%) level. The YAS-ICBT setting had significantly lower CBAY-C Modeling scores than ICBT ( $\gamma_{010} = -0.91, p = .006, d = -0.47$ ).

### Rehearsal

A linear slope was the best fitting model for the CBAY-A Rehearsal item scores, with most variability at the therapist level (96.9% intercept and 81.9% slope). As seen in Table 3, scores in the YAS-ICBT setting were significantly lower at the beginning of treatment ( $\gamma_{010} = -.73, p = .002, d = -.48$ ). There were also significant setting differences in slope ( $\gamma_{110} = -.11, p = .003, d = -1.32$ ); Rehearsal item scores in the ICBT setting increased over treatment, whereas the YAS-ICBT setting slope was essentially zero. See Fig. 4.

For CBAY-C Rehearsal item scores, the intercept only model was the best fitting model, with variance at the therapist (46.5%) and session (56.5%) level. As hypothesized, scores were

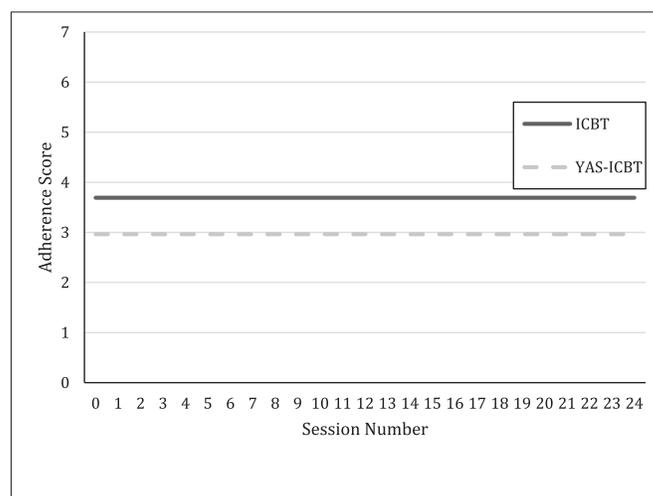


FIG. 2 Change in CBAY-A Collaborative Teaching.

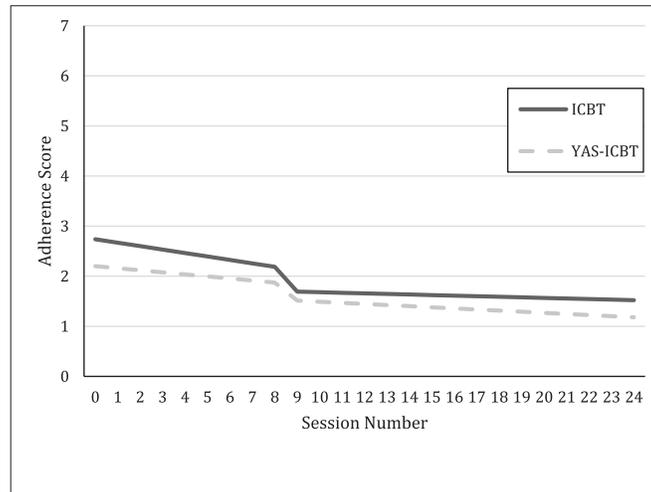


FIG. 3 Change in CBAY-A Modeling.

significantly lower in the YAS-ICBT setting than in ICBT ( $\gamma_{010} = -1.41, p < .001, d = -1.31$ ).

#### RULING OUT ALTERNATIVE INTERPRETATIONS

We examined whether findings held when controlling for youth characteristics that differed across settings (see Table 1). Control variables were entered simultaneously into each model, using multiple imputation to account for missing data. As in Table 3, most findings were unchanged by the introduction of control variables, with two exceptions. First, the setting differences in the CBAY-A Didactic Teaching item were no longer significant. To understand which control variables

might be driving this finding, we conducted post hoc analyses controlling for each control variable one at a time. In these individual analyses, the original pattern of findings held, suggesting that the change in results is associated with a combination of these variables. Second, the setting differences in the CBAY-A Rehearsal score were no longer significant. In all post hoc analyses, the setting difference in the CBAY-A Rehearsal score at the first session remained significant, suggesting that this change in results was due to a combination of the variables. The setting difference in linear slope remained significant in all models except the analysis controlling for weeks in treatment, suggesting the setting difference in

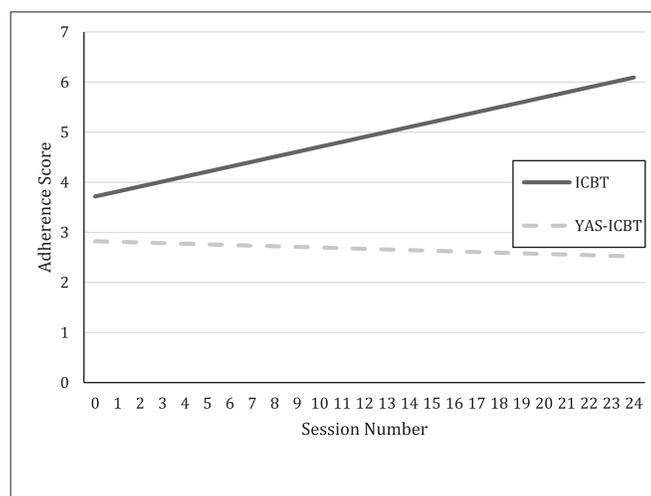


FIG. 4 Change in CBAY-A Rehearsal.

changes in Rehearsal over time may be explained by differences in the length of treatment between settings.

### Discussion

The present study examined whether the delivery of content in ICBT for youth anxiety differed across research and community settings. We found that setting did matter—for both the quantity and quality of specific ways of delivering ICBT content, community therapists scores were lower than those of research therapists. The extent to which didactic teaching, collaborative teaching, and rehearsal were used systematically varied over treatment. Differences in the quantity and quality of delivery observed between settings largely held when differences in youth characteristics between the settings were included in the model. Findings have implications for efforts to define, conceptualize, and measure treatment integrity.

Our findings underscore the potential value of separating delivery from content. Previous studies have attempted to separate delivery and content (e.g., Garland et al., 2010; Morgan et al., 2018). However, to our knowledge, this is the first study to demonstrate that the quantity and quality of delivery can be reliably assessed and used to detect differences in delivery of interventions found in an ICBT approach across settings. Our results are consistent with previous findings that suggest both the adherence and competence of content for the same EBT varies across research and community settings (McLeod et al., 2019; Smith et al., 2017). When considered with previous findings supporting the score reliability and validity of the delivery items (McLeod et al., 2018; Southam-Gerow et al., 2016), our findings provide support for the elaborative validity of the item scores (Foster & Cone, 1995). Thus, our findings indicate that it is possible to characterize the quantity and quality of how interventions commonly found in ICBT for youth anxiety are delivered.

Our findings also suggest that there were variations in the active and passive delivery of CBT content across settings. In this study, we defined didactic teaching and modeling as passive strategies that require less client involvement. Passive delivery strategies are appropriate to teach new knowledge and skills (e.g., Coleman, 1977), and are found in the first half of Coping Cat that focuses on psychoeducation (Kendall & Hedtke, 2006). A piecewise model was the best fit for the didactic teaching and modeling quantity items, suggesting that therapists utilized different quantities of passive strategies across the skill-building

and exposure phases. Both the research and community therapists used didactic teaching less during the exposure phase relative to their use in the skill-building phase, but research therapists evidenced a greater overall drop. A similar piecewise pattern was observed from research therapists for modeling, though the community therapists did not vary in their dosage of modeling across phases. Overall, the pattern of passive strategies observed in the research setting is consistent with the transition from the skill-building to the exposure phases in the Coping Cat program (Kendall & Hedtke, 2006), though the pattern of passive strategies in the community setting was less consistent.

Turning to the active strategies, research therapists delivered a higher quantity of collaborative teaching and rehearsal at the outset of treatment, but the pattern of change was not the same for the two active strategies. No change was observed over treatment for collaborative teaching. In contrast, research therapists increased their use of rehearsal over treatment, whereas no change over treatment was observed for community therapists. This is consistent with findings focused on the content of ICBT from the same RCTs in which exposure for anxiety, which requires active youth involvement, was higher in the research setting (McLeod et al., 2019). That said, though research therapists still showed increases in rehearsal over time when differences between settings were controlled, the difference between settings was no longer significant, which post hoc analyses suggested was due to controlling for differences in treatment length. These findings suggest that therapists in both settings used a consistent amount of collaborative teaching over treatment. Socratic questioning is considered a core delivery approach used in cognitive approaches (Overholser, 2011); our findings suggest that this approach was used equally by therapists in both settings. In contrast, research therapists steadily increased their use of rehearsal, which is consistent with the move from skills-training to exposure phases in Coping Cat.

Taken together, our findings suggest that the quantity and quality of delivery varies across settings. As noted earlier, the dosage of active and passive strategies over treatment in the research setting is largely consistent with the transition from the skill-building to exposure phases in Coping Cat (Kendall & Hedtke, 2006). That said, it is an open question how the quantity and quality of delivery impact treatment receipt (Fjermestad, McLeod, Tully, & Liber, 2016). Treatment receipt has been conceptualized as a combination of behavioral indicators of both client involvement (i.e., participation in therapeutic activities; Chu & Kendall, 2004) and client

comprehension (i.e., client uptake of skills; Bellg et al., 2004). Treatment receipt is a potentially important but understudied variable (see Fjermestad et al., 2016, for a review). The use of passive and active strategies may play a role in treatment receipt, particularly client comprehension. Rehearsal with feedback was identified as a superior teaching strategy over didactic teaching and modeling in a group of parents learning a behavioral parenting intervention (Jensen, Blumberg, & Browning, 2017). In a study of supervision processes in community clinics, behavioral teaching strategies (modeling, rehearsal) were associated with greater use of target practices relative to supervision that relied on didactics (Bearman et al., 2013). Future research can determine whether the use of specific passive and active strategies in different arrangements over the course of treatment can optimize client comprehension and, ultimately, client outcomes.

It is important to consider how to interpret differences in the delivery of ICBT interventions across research and community settings. On one hand, differences may represent necessary adaptations made to the delivery of the Coping Cat program given differences in youth characteristics (Ehrenreich-May et al., 2011; Southam-Gerow et al., 2008). Clinical outcomes for youth in both trials were comparable (64.0% of youth in the research setting no longer had principal diagnosis whereas 66.7% of youth in the community setting met the same criterion; Kendall et al., 2008; Southam-Gerow et al., 2010), so this interpretation warrants serious consideration. On the other hand, the differences may represent potential targets for quality improvement, such that boosting the delivery of active strategies in the community settings might further improve clinical outcomes. For example, given the importance of exposure in ICBT for anxiety (e.g., Carey, 2011)—a core component that is built around a rehearsal delivery strategy—and the low rates of exposure observed in community settings (Hipol & Deacon, 2013), the low quantity and quality of rehearsal in the community setting may represent a specific improvement target. Overall, differences in delivery across settings held when the differences between settings were controlled, with the exception of didactic teaching and rehearsal. This suggests that the variables in this study do not account for the observed differences. However, we were not able to enter a number of important variables into our models, so it is plausible that variables not measured may explain the observed differences. More work is thus needed to know what quantity and quality of delivery strategies achieve optimal treatment outcomes in either setting. Such research

may contribute to an understanding of why EBT effect sizes tend to shrink when EBTs are implemented in clinically representative settings and compared to usual clinical care (e.g., Weisz et al., 2013).

In considering our findings a few definitional issues require attention. To date, most treatment integrity instruments have focused on characterizing content and do not intentionally assess delivery. This may be a partial product of how treatment protocols are written. Few treatment protocols specify what type and amount of passive and active delivery strategies are considered appropriate at different points in the treatment process. As a result, it is hard to define delivery in terms of adherence and competence to the treatment protocol, since these protocols do not necessarily specify how interventions should be delivered from session to session. Indeed, as Kendall and Frank (2018) note, therapists should be encouraged to deliver ICBT components flexibly to meet the needs of individual youth and families. We believe that delivery item scores should be interpreted as the quantity and quality of how interventions commonly found in ICBT protocols for youth anxiety were delivered over the course of treatment, rather than as treatment integrity, and have used the terms “quantity” and “quality” rather than “adherence” and “competence” in an effort to convey that the scores do not represent protocol-prescribed delivery strategies. We are not aware of any definitions of treatment integrity that consider both delivery and content. If future research indicates that the separation of delivery and content has value, then this may have implications for how treatment integrity is defined and measured.

Relatedly, a focus on measuring treatment delivery may represent a path toward summarizing therapist activity at the case level. Treatment integrity measurement typically occurs at the session level and can be challenging to aggregate at the case level because prescribed treatment content changes session to session. For example, it can be difficult to interpret mean content item scores on the CBAY-A and CBAY-C (see Southam-Gerow et al., 2016; McLeod et al., 2019) because most Coping Cat components are designed to be delivered in one or two sessions. In contrast, delivery strategies cut across ICBT for anxiety components and may even facilitate comparisons between different CBT models (e.g., CBT for anxiety versus CBT for depression). As more modular, practice elements, and transdiagnostic approaches are developed (e.g., Ehrenreich-May, Goldstein, Wright, & Barlow, 2009; Weisz et al., 2012), a delivery-focused measurement approach

may augment traditional treatment integrity measurement for these more flexible treatment models; the prescribed treatment components may vary depending on the case, but delivery strategies are likely to generalize across cases.

A few limitations bear mentioning. First, our ability to tease apart youth and therapist effects was limited because each therapist saw few cases, especially in the community setting. Second, we were restricted in our ability to identify factors that accounted for setting-level differences. Although our analyses demonstrated that youth differences did not account for our findings, this was a quasi-experimental study and alternate explanations—including cohort effects—cannot be ruled out. Third, we could not consider therapist training and background because these variables were confounded with setting. Fourth, the two trials included Coping Cat experts in therapist training and supervision and followed the same procedures. However, adherence to the training and supervision procedures was not recorded, so it is not possible to determine if differences between the two RCTs in the quantity, quality, or differential emphasis of active versus passive delivery strategies in training or supervision influenced the findings. Despite these limitations, the present study provides preliminary evidence that the way in which ICBT for child anxiety was delivered differed between community and research settings.

## References

- Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington: University of Vermont, Department of Psychiatry.
- Barber, J. P., Sharpless, B. A., Klostermann, S., & McCarthy, K. S. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology: Research and Practice*, 38, 493–500. <https://doi.org/10.1037/0735-7028.38.5.493>
- Bearman, S., Weisz, J. R., Chorpita, B. F., Hoagwood, K., Ward, A., Ugueto, A. M., . . . Research Network on Youth Mental Health (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. <https://doi.org/10.1007/s10488-013-0485-5>
- 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. [https://doi.org/10.1016/S0005-7967\(03\)00138-4](https://doi.org/10.1016/S0005-7967(03)00138-4)
- Bellg, 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. <https://doi.org/10.1037/0278-6133.23.5.443>
- Bornstein, M. R., Bellack, A. S., & Hersen, M. (1977). Social-skills training for unassertive children: A multiple baseline analysis. *Journal of Applied Behavior Analysis*, 10, 183–195. <https://doi.org/10.1901/jaba.1977.10-183>
- Carey, T. A. (2011). Exposure and reorganization: The what and how of effective psychotherapy. *Clinical Psychology Review*, 31, 236–248. <https://doi.org/10.1016/j.cpr.2010.04.004>
- Chu, B. C., & Kendall, P. C. (2004). Positive association of child involvement and treatment outcome within a manual-based Cognitive-Behavioral Treatment for children with anxiety. *Journal of Consulting and Clinical Psychology*, 72, 821–829. <https://doi.org/10.1037/0022-006X.72.5.821>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Coleman, J. A. (1977). Differences between experiential and classroom learning. In M. T. Keeton (Ed.), *Experiential learning: Rationale, characteristics, and assessment* (pp. 49–61). San Francisco, CA: Jossey-Bass.
- Deacon, B. J., Farrell, N. R., Kemp, J. J., Dixon, L. J., Sy, J. T., Zhang, A. R., & McGrath, P. B. (2013). Assessing therapist reservations about exposure therapy for anxiety disorders: The Therapist Beliefs about Exposure Scale. *Journal of Anxiety Disorders*, 27, 772–780. <https://doi.org/10.1016/j.janxdis.2013.04.006>
- Ehrenreich-May, J. T., Goldstein, C. R., Wright, L. R., & Barlow, D. H. (2009). Development of a unified protocol for the treatment of emotional disorders in youth. *Child and Family Behavioral Therapy*, 31, 20–37. <https://doi.org/10.1080/07317100802701228>
- Ehrenreich-May, J., Southam-Gerow, M. A., Hourigan, S. E., Wright, L. R., Pincus, D. B., & Weisz, J. R. (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, 398–411. <https://doi.org/10.1007/s10488-010-0328-6>
- Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods*, 14, 43–53. <https://doi.org/10.1037/a0014699>
- Fjermestad, K. W., McLeod, B. D., Tully, C. B., & Liber, J. M. (2016). Therapist characteristics and interventions: Enhancing alliance and involvement with youth. In S. Maltzman (Ed.), *Oxford handbook of treatment processes and outcomes in counseling psychology* New York: Oxford University Press
- Foster, S. L., & Cone, J. D. (1995). Validity issues in clinical assessment. *Psychological Assessment*, 7, 248–260. <https://doi.org/10.1037/1040-3590.7.3.248>
- 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. <https://doi.org/10.1176/ps.2010.61.8.788>
- Hipol, L. J., & Deacon, B. J. (2013). Dissemination of evidence-based practices for anxiety disorders in Wyoming: A survey of practicing psychotherapists. *Behavior Modification*, 37, 170–188. <https://doi.org/10.1177/0145445512458794>
- IBM Corp (2017). *SPSS Statistics for Windows*. Armonk, NY: IBM Corp.
- Jensen, S. A., Blumberg, S., & Browning, M. (2017). Structured feedback training for time-out: Efficacy and efficiency in

- comparison to a didactic method. *Behavior Modification*, 42, 765–780. <https://doi.org/10.1177/0145445517733474>
- Kendall, P. C. (1994). Treating anxiety disorders in children: Results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 62(1), 100–110. <https://doi.org/10.1037/0022-006X.62.1.100>
- Kendall, P. C., Flannery-Schroeder, E., Panichelli-Mindel, S., Southam-Gerow, M. A., Henin, A., & Warman, M. (1997). Therapy for youths with anxiety disorders: A second randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 65(3), 366–380. <https://doi.org/10.1037/0022-006X.65.3.366>
- Kendall, P. C., & Frank, H. E. (2018). Implementing evidence-based treatment protocols: Flexibility within fidelity. *Clinical Psychology: Science and Practice*, 25(4). <https://doi.org/10.1111/cpsp.12271>
- 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., & Suveg, C. (2008). Cognitive-behavioral therapy for anxiety disorder youth: A randomized clinical trial evaluating child and family modalities. *Journal of Consulting and Clinical Psychology*, 76(2), 282–297. <https://doi.org/10.1037/0022-006X.76.2.282>
- Kolb, D. A., Boyatzis, R. E., & Mainemelis, C. (2000). Experiential learning theory: Previous research and new directions. In R. J. Sternberg, & L. F. Zhang (Eds.), *Perspectives on cognitive, learning, and thinking styles* NY: Routledge
- Lyon, A. R., & Koerner, K. (2016). User-centered design for psychosocial intervention development and implementation. *Clinical Psychology: Science and Practice*, 23(2), 180–200. <https://doi.org/10.1111/cpsp.12154>
- McLeod, B. D., Southam-Gerow, M. A., Jensen-Doss, A., Hogue, A., Kendall, P. C., & Weisz, J. R. (2019). Benchmarking treatment adherence and therapist competence in individual cognitive-behavioral treatment for youth anxiety disorders. *Journal of Clinical Child and Adolescent Psychology*, 48(S1), S234–S246. <https://doi.org/10.1080/15374416.2017.1381914>
- McLeod, B. D., Southam-Gerow, M. A., Rodriguez, A., Quinoy, A., Arnold, C., Kendall, P. C., & Weisz, J. R. (2018). Development and initial psychometrics for a therapist competence instrument for CBT for youth anxiety. *Journal of Clinical Child and Adolescent Psychology*, 47(1), 47–60. <https://doi.org/10.1080/15374416.2016.1253018>
- Morgan, N. R., Davis, K. D., Richardson, C., & Perkins, D. F. (2018). Common components analysis: An adapted approach or evaluating programs. *Evaluation and Program Planning*, 67, 1–9. <https://doi.org/10.1016/j.evalprogplan.2017.10.009>
- Overholser, J. C. (2011). Collaborative empiricism, guided discovery, and the Socratic method: Core processes for effective cognitive therapy. *Clinical Psychology: Science and Practice*, 18(1), 62–66. <https://doi.org/10.1111/j.1468-2850.2011.01235.x>
- Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice*, 12, 365–383. <https://doi.org/10.1093/clipsy/bpi045>
- 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.
- Ross, A. O. (1964). Learning theory and therapy with children. *Psychotherapy: Theory, Research & Practice*, 1, 102–108. <https://doi.org/10.1037/h0088580>
- Shaffer, D., Fisher, P., Lucas, C., Dulcan, M. K., & Schwab-Stone, M. (2000). NIMH Diagnostic Interview Schedule for Children, Version IV: Description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry*, 39, 28–38. <https://doi.org/10.1097/00004583-200001000-00014>
- 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. <https://doi.org/10.1037/0022-006X.73.1.106>
- Silverman, W. K., & Albano, A. M. (1996). *The Anxiety Disorders Interview Schedule for Children for DSM-IV: (Child and Parent Versions)*. San Antonio, TX: Psychological Corporation.
- Smith, M. M., McLeod, B. D., Southam-Gerow, M. A., Jensen-Doss, A., Kendall, P. C., & Weisz, J. R. (2017). Does the delivery of CBT for youth anxiety differ across research and practice settings? *Behavior Therapy*, 48, 501–516. <https://doi.org/10.1016/j.beth.2016.07.004>
- Southam-Gerow, M. A., Chorpita, B. F., Miller, L. M., & Gleacher, A. A. (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. <https://doi.org/10.1007/s10488-007-0154-7>
- Southam-Gerow, M. A., McLeod, B. D., Arnold, C. C., 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, 70–80. <https://doi.org/10.1037/pas0000141>
- Southam-Gerow, M. A., Weisz, J. R., Chu, B. C., McLeod, B. D., Gordis, E. B., & Connor-Smith, J. K. (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, 1043–1052. <https://doi.org/10.1016/j.jaac.2010.06.009>
- Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S. K., Miranda, J., Bearman, S. K., . . . Research Network on Youth Mental Health (2012). Testing standard and modular designs for psychotherapy treating depression, anxiety, and conduct problems for youth: A randomized effectiveness trial. *Archives of General Psychiatry*, 69, 274–282. <https://doi.org/10.1001/archgenpsychiatry.2011.147>
- Weisz, J. R., Krumholz, L. S., Santucci, L., Thomassin, K., & Ng, M. (2015). Shrinking the gap between research and practice: Tailoring and testing youth psychotherapies in clinical care contexts. *Annual Review of Clinical Psychology*, 11, 139–163. <https://doi.org/10.1146/annurev-clinpsy-032814-112820>
- 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, 750–761. <https://doi.org/10.1001/jamapsychiatry.2013.1176>

RECEIVED: July 17, 2019  
ACCEPTED: November 19, 2019  
AVAILABLE ONLINE: XXXX