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.

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 continuum 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 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 treatment 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 ethnic/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 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.

ASSSESSMENTS 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 naive 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, ICBT = 0):

\[
L1: \text{CBAYADITCH}_{ijk} = \pi_{0ijk} + \pi_{1jk}(\text{TIMEWEEK}_{ijk}) + \pi_{2jk}(\text{EXPCHG}_{ijk}) + \pi_{3jk}(\text{EXPOSURE}_{ijk}) + e_{ijk}
\]

\[
L2: \pi_{0ijk} = \beta_{00k} + \beta_{01k}(\text{YAS-ICBT}_{jk}) + \tau_{0jk}
\]

\[
\pi_{1jk} = \beta_{10k} + \beta_{11k}(\text{YAS-ICBT}_{jk}) + \tau_{1jk}
\]

\[
\pi_{2jk} = \beta_{20k} + \beta_{21k}(\text{YAS-ICBT}_{jk}) + \tau_{2jk}
\]

\[
\pi_{3jk} = \beta_{30k} + \beta_{31k}(\text{YAS-ICBT}_{jk}) + \tau_{3jk}
\]

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

\[
\beta_{01k} = \gamma_{010}
\]

\[
\beta_{10k} = \gamma_{100} + u_{10k}
\]

\[
\beta_{11k} = \gamma_{110}
\]

\[
\beta_{20k} = \gamma_{200} + u_{20k}
\]

\[
\beta_{21k} = \gamma_{210}
\]

\[
\beta_{30k} = \gamma_{300} + u_{30k}
\]

\[
\beta_{31k} = \gamma_{310}
\]

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 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.

<table>
<thead>
<tr>
<th>Variable</th>
<th>ICBT (%)</th>
<th>YAS-ICBT (%)</th>
<th>$F$ or $P$</th>
<th>Chi Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>10.36</td>
<td>11.32</td>
<td>2.94</td>
<td>.091</td>
</tr>
<tr>
<td>(1.90)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>60.8</td>
<td>29.4</td>
<td>5.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td>White</td>
<td>86.3</td>
<td>41.2</td>
<td>15.48</td>
</tr>
<tr>
<td></td>
<td>African-American</td>
<td>9.80</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Latino</td>
<td>1.9</td>
<td>17.6</td>
<td>4.70</td>
</tr>
<tr>
<td></td>
<td>Mixed/Other</td>
<td>1.9</td>
<td>5.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Not Reported</td>
<td>-</td>
<td>35.3</td>
<td>22.81</td>
</tr>
<tr>
<td>CBCL</td>
<td>Total</td>
<td>62.92</td>
<td>64.19</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(8.56)</td>
<td>(7.34)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Internalizing</td>
<td>67.08</td>
<td>66.38</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(8.60)</td>
<td>(8.33)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Externalizing</td>
<td>52.76</td>
<td>60.81</td>
<td>8.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(10.08)</td>
<td>(7.49)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anxious-Depressed</td>
<td>62.94</td>
<td>68.63</td>
<td>4.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(9.29)</td>
<td>(8.69)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Principal Diagnoses</td>
<td>22.81</td>
<td>22.81</td>
<td>22.81</td>
</tr>
<tr>
<td></td>
<td>GAD</td>
<td>37.3</td>
<td>5.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SAD</td>
<td>29.4</td>
<td>35.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SOP</td>
<td>33.3</td>
<td>23.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SP</td>
<td>-</td>
<td>35.3</td>
<td>35.3</td>
</tr>
<tr>
<td></td>
<td>Family Income</td>
<td>Up to 60k</td>
<td>35.3</td>
<td>70.6</td>
</tr>
<tr>
<td></td>
<td>Number of Sessions</td>
<td>15.92</td>
<td>16.82</td>
<td>1.36</td>
</tr>
<tr>
<td></td>
<td>(1.43)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weeks in Treatment</td>
<td>19.52</td>
<td>26.38</td>
<td>15.67</td>
</tr>
<tr>
<td></td>
<td>(3.97)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of Coded Sessions</td>
<td>10.43</td>
<td>12.47</td>
<td>4.71</td>
</tr>
<tr>
<td></td>
<td>(2.84)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.61)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Note: ICBT &gt; YAS-ICBT</th>
</tr>
</thead>
<tbody>
<tr>
<td>a = YAS-ICBT &gt; ICBT</td>
</tr>
</tbody>
</table>

Please cite this article as: J. R. Cox, B. D. McLeod, , et al., Examining How CBT Interventions for Anxious Youth Are Delivered Across Settings, Behavior Therapy, https://doi.org/10.1016/j.beth.2019.11.008
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 \( \chi^2 = 236.8, DF = 237, p = .49 \)). 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 (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 level (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
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 ($\gamma_{010} = -0.73$, $p = .002$, $d = -0.48$) 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
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
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
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 impor-
tant 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 comprehen-
sion 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 adap-
tations 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 (Hipop & Deacon, 2013), the
low quantity and quality of rehearsal in the
community setting may represent a specific im-
provement 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 size trends tend to shrink when EBTs are imple-
mented 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 characteriz-
ing 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

Please cite this article as: J. R. Cox, B. D. McLeod, , et al., Examining How CBT Interventions for Anxious Youth Are Delivered Across
Settings, Behavior Therapy, https://doi.org/10.1016/j.beth.2019.11.008
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


**Delivered Strategies in ICBT for Anxious Youth**

To cite this article as: J. R. Cox, B. D. McLeod, . . . et al., Examining How CBT Interventions for Anxious Youth Are Delivered Across Settings, *Behavior Therapy*, https://doi.org/10.1016/j.beth.2019.11.008

**Please cite this article as:** J. R. Cox, B. D. McLeod, . . . et al., Examining How CBT Interventions for Anxious Youth Are Delivered Across Settings, *Behavior Therapy*, https://doi.org/10.1016/j.beth.2019.11.008

[108.162.210.203]