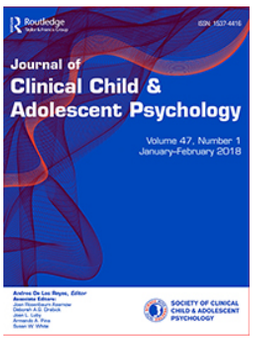

Routledge
Taylor & Francis Group

Journal of Clinical Child & Adolescent Psychology

ISSN: 1537-4416 (Print) 1537-4424 (Online) Journal homepage: http://www.tandfonline.com/loi/hcap20

Development and Initial Psychometrics for a Therapist Competence Instrument for CBT for Youth Anxiety

Bryce D. McLeod, Michael A. Southam-Gerow, Adriana Rodríguez, Alexis M. Quinoy, Cassidy C. Arnold, Philip C. Kendall & John R. Weisz

To cite this article: Bryce D. McLeod, Michael A. Southam-Gerow, Adriana Rodríguez, Alexis M. Quinoy, Cassidy C. Arnold, Philip C. Kendall & John R. Weisz (2018) Development and Initial Psychometrics for a Therapist Competence Instrument for CBT for Youth Anxiety, Journal of Clinical Child & Adolescent Psychology, 47:1, 47-60, DOI: 10.1080/15374416.2016.1253018

To link to this article: https://doi.org/10.1080/15374416.2016.1253018

Journal of Clinical Child & Adolescent Psychology, 47(1), 47–60, 2018

ISSN: 1537-4416 print/1537-4424 online
DOI: 10.1080/15374416.2016.1253018

Routledge
Taylor & Francis Group

Development and Initial Psychometrics for a Therapist Competence Instrument for CBT for Youth Anxiety

Bryce D. McLeod, Michael A. Southam-Gerow, Adriana Rodríguez, Alexis M. Quinoy, and Cassidy C. Arnold

Department of Psychology, Virginia Commonwealth University

Philip C. Kendall

Department of Psychology, Temple University

John R. Weisz

Department of Psychology, Harvard University

Therapist competence is an important component of treatment integrity. This article reports on the development and initial psychometric assessment of the Cognitive-Behavioral Treatment for Anxiety in Youth Competence Scale (CBAY-C), an observational instrument designed to capture therapist limited-domain competence (i.e., competence in the delivery of core interventions and delivery methods found in a specific psychosocial treatment program) in the delivery of the core practice elements in individual cognitive-behavioral treatment (ICBT) for youth anxiety. Treatment sessions ($N = 744$) from 68 youth participants (M age = 10.60 years, $SD = 2.03$; 82.3% Caucasian; 52.9% male) of the same ICBT program for youth anxiety from (a) an efficacy study and (b) an effectiveness study were independently scored by 4 coders using observational instruments designed to assess therapist competence, treatment adherence, treatment differentiation, alliance, and client involvement. Interrater reliability—intraclass correlation coefficients (2,2)—for the item scores averaged 0.69 ($SD = 0.11$). The CBAY-C item, scale, and subscale (Skills, Exposure) scores showed evidence of validity via associations with observational instruments of treatment adherence to ICBT for youth anxiety, theory-based domains (cognitive-behavioral treatment, psychodynamic, family, client centered), alliance, and client involvement. Important to note, although the CBAY-C scale, subscale, and item scores did overlap with a corresponding observational treatment adherence instrument independently rated by coders, the degree of overlap was moderate, indicating that the CBAY-C assesses a distinct component of treatment integrity. Applications of the instrument and future research directions discussed include the measurement of treatment integrity and testing integrity-outcome relations.

Competence in the delivery of therapeutic interventions is considered a key part of successful psychosocial treatment (Barber, Sharpless, Klostermann, & McCarthy, 2007). Despite this widely held contention, measurement of therapist competence is underdeveloped, particularly in treatment for children and adolescents (Webb, Derubeis, & Barber, 2010). Thus, experts and federal funding agencies have called for efforts to assess therapist competence (Barber et al., 2007) as a variable associated with differential client outcomes.

Therapist competence is generally considered a component of a broader concept, treatment integrity (Perepletchikova & Kazdin, 2005). Treatment integrity refers to the degree to which a treatment was delivered as intended and is composed of *adherence* (carrying out procedures called for in the model), *differentiation* (*not* carrying out procedures that are proscribed in the model), and

Correspondence should be addressed to Bryce D. McLeod, Department of Psychology, Virginia Commonwealth University, 806 W. Franklin Street, P.O. Box 842018, Richmond, VA 23284-2018. E-mail: bmcleod@vcu.edu

competence (the quality of treatment delivery; Perepletchikova & Kazdin, 2005). Reviews have concluded that few clinical trials have assessed all components (Goense, Boendermaker, Van Yperen, Stams, & Van Laar, 2014; Perepletchikova, Treat, & Kazdin, 2007). Of the integrity components, adherence has received most attention, but competence is understudied in the child and adolescent (hereafter, youth) treatment field (Goense et al., 2014; Perepletchikova et al., 2007).

Definitional issues have, in part, slowed efforts to study therapist competence. Broadly speaking, definitions of therapist competence fall into two categories. The first focuses on general skills believed to apply across treatment domains. A variety of terms are used for this category, including common-factors (e.g., Castonguay, 1993), foundational (Rodolfa et al., 2005), and global (Barber et al., 2007) competency. This broad category encompasses skillfulness and responsiveness in delivering specific therapeutic interventions along with a therapist's ability to develop and maintain a strong alliance with a client. It is hypothesized that common competence may increase the effectiveness of treatment via the formation of a positive alliance characterized by a strong bond and agreement on therapeutic activities and goals (Barber et al., 2007). The second category focuses on the delivery of specific interventions (e.g., exposures) and delivery methods (e.g., rehearsal, coaching) found in treatment programs; this has been called limited-domain competence (also technical and model-specific competence; Barber et al., 2007). Limited domain competence is a narrower construct and is often considered a subcategory of global competence (Barber et al., 2007). Although both categories are considered important, only limited domain competence has been the focus of empirical attention (Webb et al., 2010).

To our knowledge, only five observational instruments designed to assess limited-domain competence (hereafter, competence) exist in the youth treatment field (Competence and Adherence Scale for Cognitive Behavioral Therapy: Bjaastad et al., 2016; Fidelity of Implementation Rating System: Forgatch, Patterson, & Degarmo, 2005; Therapeutic Competence Scale: Gutermann et al., 2014; Therapist Behavior Rating Scale-Competence: Hogue et al., 2008a; Cognitive Behaviour Therapy Scale for Children and Young People: Stallard, Myles, & Branson, 2014). Four of these gauge competence in delivering a specific manualized treatment (Bjaastad et al., 2016; Forgatch et al., 2005; Gutermann et al., 2014; Hogue et al., 2008a), and one assesses competencies associated with delivering cognitive-behavioral treatment (CBT; Stallard et al., 2014). Scores on four of the instruments have demonstrated low to strong interrater item reliability calculated with Kappa ($\kappa >$ .69; Forgatch et al., 2005) and intraclass correlation coefficients (ICCs; .01–.87; Bjaastad et al., 2016; Gutermann et al., 2014; Hogue et al., 2008a); Stallard et al. did not calculate interrater reliability, as the coders discussed items and generated a consensus score. Analyses indicated that the scores on three instruments demonstrated convergent validity with adherence (Bjaastad et al., 2016; Hogue et al., 2008a; Stallard et al., 2014), two demonstrated discriminant validity (Hogue et al., 2008a; Stallard et al., 2014), and two evidenced predictive validity for clinical outcomes (Forgatch et al., 2005; Hogue et al., 2008b).

There are differences among the existing instruments. First, they differ in item focus. Some of the instruments emphasized global ratings of competence (i.e., overall quality of delivery of a particular treatment model, like CBT; e.g., Stallard et al., 2014); others assess competence in more granular, model-specified interventions (e.g., exposures in PTSD treatment); and still others assess competence in the *method* of delivering content (e.g., Socratic questioning; e.g., Forgatch et al., 2005). The variation in focus underscores the fact that these are early days in measuring competence in youth treatment, with little consensus yet on which aspects of the treatment process should be assessed.

Another difference is that one of the five instruments assessed "competent adherence" that combines adherence and competence in a single rating (Forgatch et al., 2005), whereas the other instruments focus solely on competence (e.g., Stallard et al., 2014) or produce separate adherence and competence ratings (e.g., Bjaastad et al., 2016). With regard to scoring strategy, all instruments rely on Likert-type ratings.

These differences revealed important foci for future research that we sought to address. The first area concerns how limited-domain competence is conceptualized: as a global or discrete phenomenon (or both; Hogue et al., 2008a). Global approaches use one item (e.g., How competent was the therapist in delivering CBT for youth anxiety?) or a few (e.g., How responsive was the therapist when delivering CBT for youth anxiety? How skillful was the therapist when delivering CBT for youth anxiety?) to assess overall quality in delivering a particular treatment model. A potential advantage of this approach is efficiency (e.g., time/money for coders). In contrast, using discrete items involves separate ratings for the core theory-driven interventions and delivery methods found in a treatment program (e.g., Carroll et al., 2000; Hogue et al., 2008a). An advantage of this approach is that the competence of each core intervention and delivery method is assessed, allowing for tests of theories of change. Given the potential advantages of both approaches, and the fact that previous research has not examined the correspondence between the approaches, we developed an instrument that includes both global and discrete competence ratings.

A second area we sought to address concerns which aspects of CBT for youth anxiety should be assessed with discrete items. Our goal was to develop an instrument that could capture the core components of individual-CBT (ICBT) for youth anxiety. To accomplish this aim, we

identified three categories of items. First, we included items that assess competence in delivering interventions common to many CBT approaches (e.g., homework assignment; Stallard et al., 2014). Second, and consistent with most competence instruments (Bjaastad et al., 2016; Forgatch et al., 2005; Gutermann et al., 2014; Hogue et al., 2008a), we developed items that gauge competence in delivering the core theory-driven interventions central to ICBT for youth anxiety (e.g., coping plan, exposure). Third, we developed items that measure competence in specific delivery methods (e.g., didactically or via rehearsal; Southam-Gerow et al., 2016), an approach found in a previous competence instrument (Forgatch et al., 2005).

The final question we addressed was the overlap between adherence and competence. Conceptually, the two are distinct (Perepletchikova & Kazdin, 2005); however, the degree of overlap has varied from acceptable to unacceptable levels, with correlations ranging from 0.31 (Carroll et al., 2000) to 0.96 (Barber, Liese, & Abrams, 2003). Unfortunately, common-method bias—using the same coders to rate both adherence and competence—may have contributed to inflated correlations in past studies. However, it is also possible that adherence and competence are not distinguishable. For integrity instruments to be useful they need to discriminate between the quantity and quality in the delivery of core interventions and delivery methods from specific manualized treatments (Hogue et al., 2008a). Thus, critical to establishing the representative validity (i.e., determining if the instrument assesses what it purports to measure; Foster & Cone, 1995) of a competence instrument is to ensure that it is differentiated from adherence ratings for the same manualized treatment.

This study reports on the development and initial psychometric evaluation of scores on the Cognitive-Behavioral Treatment for Anxiety in Youth Competence Scale (CBAY-C), an instrument designed to assess therapist competence in the delivery of ICBT for youth anxiety. Our purpose was to develop an instrument that could be used to support the implementation and evaluation of treatments (Garland & Schoenwald, 2013). We thus developed items that (a) capture practice elements (i.e., "discrete clinical technique used as part of a larger intervention plan"; Chorpita & Daleiden, 2009, p. 560) of the core theory-driven interventions found across ICBT programs and (b) gauge *how* the practice elements were delivered. These features enhance the flexibility of treatment integrity instruments and thus are preferable when conducting implementation research in diverse settings (McLeod, Southam-Gerow, Bair, Rodriguez, & Smith, 2013). Once developed, the CBAY-C was used to code sessions for youth diagnosed with anxiety disorders from the same manual-based ICBT program delivered in research and practice settings. To assess the score reliability and validity of the CBAY-C, we used a multicomponent measurement model that included treatment integrity (adherence, differentiation, competence) and relational factors (alliance, client involvement). To determine whether the CBAY-C assesses competence independently of adherence, we included the CBT Adherence Scale for Youth Anxiety (CBAY-A; Southam-Gerow et al., 2016). Because the CBAY-C and CBAY-A were designed to be parallel instruments, our approach provides a test of whether competence and adherence can be distinguished for the same treatment program (Hogue et al., 2008a).

To establish representative validity, we hypothesize that the scale, subscale, and item scores would demonstrate convergence and divergence with related and distinct treatment process domains (adherence, differentiation, alliance, client involvement) consistent with patterns of associations seen in previous studies (e.g., Carroll et al., 2000; Hogue et al., 2008a). We hypothesized that the CBAY-C scores would evidence convergent validity by demonstrating the strongest relations with the CBAY-A (Carroll et al., 2000; Hogue et al., 2008a) followed by an observational instrument designed to assess CBT interventions. We hypothesized that the CBAY-C scores would evidence discriminant validity by demonstrating small to moderate correlations with an observational instrument of interventions from non-CBT domains (Client-Centered, Psychodynamic, Family; Carroll et al., 2000) and observational instruments that assess the relational components of treatment (alliance, client involvement; Carroll et al., 2000; Hogue et al., 2008a). Finally, as client and therapist characteristics may influence competence (Barber, Foltz, Crits-Christoph, & Chittams, 2004; Boswell et al., 2013), we evaluated whether targets of measurement that might influence competence (youth, therapist, study group, coder, time in treatment) accounted for systematic variation in CBAY-C scores. We compared the pattern of variation in the CBAY-C scores to that found in the CBAY-A. If different targets of measurement account for diverse proportions of variance across the competence and adherence instruments, this would provide evidence that they assess distinct integrity components and inform future research applications of the instrument.

METHOD

Data Sources and Participants

Treatment data were collected from 68 youth who participated in two randomized controlled trials (RCTs). Kendall, Hudson, Gosch, Flannery-Schroeder, and Suveg (2008) compared the efficacy of ICBT, family-CBT, and an active control group. Southam-Gerow et al. (2010) compared the effectiveness of ICBT (YAS-ICBT) to usual care in the Youth Anxiety Study (YAS). Only the ICBT groups (ICBT, YAS-ICBT) were used in the present study. Recorded sessions collected in each RCT served as our data. To be included in this study, youth had (a) at least two audible recorded sessions and (b) received ICBT from a

single therapist. See Kendall et al. and Southam-Gerow et al. for more details about participant recruitment procedures.

The 68 youth participants 7–15 years of age (*M* age = 10.60 years, *SD* = 2.03; 82.3% Caucasian; 52.9% male) met diagnostic criteria for a primary anxiety disorder (see Table 1). Fifty-one youth participants were from Kendall et al. (2008) and 17 were from YAS (Southam-Gerow et al., 2010). In Kendall et al., 64.0% of the youth who received ICBT no longer met diagnostic criteria for their principal anxiety disorder at posttreatment. In YAS-ICBT, 66.7% of youth no longer met diagnostic criteria at posttreatment.

There were 29 therapists (69.7% Caucasian, 13.2% male). In Kendall et al. (2008), clinical psychology doctoral trainees and licensed psychologists delivered ICBT (*N* = 16, 12.5% male). The therapists were 81.1% Caucasian, 6.3% Latino, 6.3% Asian/Pacific Islander, and 6.3% did not report. Therapists in YAS-ICBT were clinic employees who volunteered to participate and were randomly assigned to YAS-ICBT or usual care. YAS-ICBT therapists (*N* = 13, 15.4% male) were 53.8% Caucasian, 15.4% Latino, 15.4% Asian/Pacific Islander, and 15.4% mixed/other. The professional makeup was 30.8% social workers, 46.1% psychologists (with 15.4% at the doctoral level), and 23.1% "other."

TABLE 1
Youth Descriptive Data and Group Comparisons

	M (SD) *or* %		
Variable	*ICBT*[a]	*YAS-ICBT*[b]	F *or* χ^2
Age	10.36 (1.90)	11.32 (2.32)	1.72
Sex			
Male	60.80	29.40	5.04*
Race/Ethnicity			18.97**
Caucasian	86.20[c]	41.20	
African American	9.80	—	
Latino	2.00	17.60[d]	
Mixed/Other	2.00	5.90	
Not Reported	—	35.30[d]	
CBCL			
Total	63.18 (8.44)	64.19 (7.34)	0.43
Internalizing	67.40 (8.37)	66.38 (8.33)	0.43
Externalizing	52.96 (10.08)	60.81 (7.49)[d]	2.87**
Primary Diagnoses			22.81**
GAD	37.30[c]	5.90	
SAD	29.40	35.30	
SOP	33.30	23.50	
SP	—	35.30[d]	
Family Income			7.92**
Up to 60k per Year	35.30	70.60[d]	
No. of Sessions	15.92 (1.43)	16.82 (5.02)	0.73
Weeks in Treatment	19.52 (3.97)	26.38 (10.41)[d]	2.65**

Note: Analysis of variance was conducted with continuous variables, whereas chi-square analyses were conducted with continuous variables. ICBT = individual cognitive-behavioral treatment delivered in Kendall et al. (2008) study; YAS-ICBT = individual cognitive-behavioral treatment delivered in YAS; CBCL = Child Behavior Checklist; GAD = generalized anxiety disorder; SAD = separation anxiety disorder; SOP = social phobia; SP = specific phobia.

[a] $N = 51$.

[b] $N = 17$.

[c] ICBT > YAS-ICBT

[d] YAS-ICBT > ICBT

*$p < .05$. **$p < .01$.

Individual CBT

Therapists in ICBT and YAS-ICBT delivered the Coping Cat program, an ICBT program for youth diagnosed with anxiety disorders (Kendall & Hedtke, 2006) that consists of 16 sessions; 14 sessions are conducted individually with the youth, and two sessions are conducted with the parents. The first half of treatment focuses on anxiety management skills training (e.g., relaxation, problem solving), whereas the second half emphasizes exposures. Homework is regularly assigned to the youth throughout the program. In both studies, quality control methods (Garland & Schoenwald, 2013) included a treatment manual, a training workshop, and weekly supervision with an expert in ICBT for youth anxiety. Adherence was measured with the Coping Cat Brief Adherence Scale (see Kendall, 1994); Coping Cat therapists in both studies showed more than 90.0% adherence (see Kendall et al., 2008; Southam-Gerow et al., 2010).

Competence Instrument Development

The CBAY-C was designed to assess competence of practice elements for ICBT for youth anxiety. The CBAY-C was modeled after exemplar observational competence instruments that had parallel adherence instruments (see, e.g., Carroll et al., 2000; Hogue et al., 2008a). The CBAY-C was developed via a four-step process: (a) scale development, wherein conceptual dimensions were identified; (b) item development, wherein items that map onto the dimensions were identified from multiple sources to establish content validity; (c) scoring strategy, wherein a scoring strategy was determined for the items; and (d) scoring manual development and pilot coding.

Scale Development

Scale categories were developed based on existing competence instruments (e.g., Carroll et al., 2000; Hogue et al., 2008a), conceptually, and with consultation from developers of ICBT approaches. Specifically, we identified four categories. The first category was Standard Interventions (e.g., Homework Review; see Table 1), which focuses on interventions that are common to CBT programs (i.e., not unique to ICBT for youth anxiety) and expected in each session (Barber et al., 2003). The second category was Model Interventions, which focuses on the core theory-driven interventions specific to ICBT for youth anxiety and expected to

be the focus of one or more sessions, such as Relaxation or Exposure (Carroll et al., 2000; Hogue et al., 2008a). Third, we identified a Delivery category, which represents *how* specific model interventions were delivered (e.g., Modeling, Rehearsal). Finally, we included two global items that focus on the overall skillfulness ("What was the overall level of therapist skillfulness in delivering ICBT for youth anxiety?") and responsiveness ("What was the overall level of therapist responsiveness in delivering ICBT for youth anxiety?") demonstrated by the therapist when delivering ICBT for youth anxiety in a session (e.g., Carroll et al., 2000).

Item Generation and Refinement

To help establish content validity, we used three sources to develop items for the scales. Our first source was the Coping Cat manual (Kendall & Hedtke, 2006) and the Modular Approach for Treating Childhood Problems manual (Chorpita & Weisz, 2009). We also reviewed the adherence instruments used in the RCTs (see Kendall et al., 2008; Weisz, Chorpita, Palinkas, Schoenwald, & Miranda, 2012). Content pulled from these sources was listed, and similar interventions and delivery approaches found in both programs were combined into a single item. The next source was a review by Chorpita and Daleiden (2009) that identified common interventions and delivery methods used in ICBT approaches for youth anxiety. Chorpita and Daleiden identified specific evidence-based treatment programs (N = 56) and then distilled the interventions contained within the programs resulting in a list of "practice elements" represented in ICBT programs for youth anxiety. The practice elements identified in the Chorpita and Daleiden review were combined with those produced from our review of the adherence instruments and treatment manuals to produce an initial list of items (N = 23). Once generated, the initial item list was distributed to experts in ICBT for youth anxiety. The experts reviewed the items and recommended that we separate the Exposure item into three items: Exposure Preparation, Exposure, and Exposure Debrief. Once this revision was made, the experts approved the final set of items. These steps resulted in an instrument with 25 total items: five Standard, 12 Model, six Delivery, and two Global (see Table 2).

Scoring Strategy

We adopted a scoring strategy for competence that includes both the *quality of delivery* (skillfulness) and the *timing and appropriateness of delivery for a given client and situation* (responsiveness). For each item, coders considered the extent to which a therapist demonstrated the four dimensions of skillfulness and responsiveness in a session (Carroll et al., 2000): (a) expertise, commitment, motivation; (b) clarity of communication; (c) appropriate timing of delivery; and (d) read and respond to where the client appears to be regarding level of therapeutic engagement/understanding. Coders were asked to consider these dimensions of skillfulness and responsiveness for each item when making ratings on a 7-point Likert-type scale with the following anchors: 1 (*very poor*), 3 (*acceptable*), 5 (*good*), and 7 (*excellent*). This scoring strategy is used in exemplar competence coding systems developed for youth (Hogue et al., 2008a) and adult (Barber et al., 2003; Carroll et al., 2000) treatment. Competence ratings were made only when the item occurred in a session.

Scoring Manual Development

Following these steps, a preliminary scoring manual was developed and pilot coding was used to refine the scoring manual. Following completion of the pilot coding, a final version of the scoring manual was produced (Southam-Gerow et al., 2010), which provides coders with a comprehensive guide for coding sessions. The manual contains a thorough description of each item and provides additional information to help the coder make coding decisions in an informed and reliable manner.

Other Instruments Used for Validity Analyses

Cognitive Behavioral Treatment Adherence Scale for Youth Anxiety (CBAY-A; Southam-Gerow et al., 2016) is a 22-item instrument designed to assess treatment adherence for ICBT for youth anxiety with a structure designed to parallel the CBAY-C (Table 3 lists overlapping items). The items measure three areas: (a) *Standard*, four items that represent standard CBT interventions (e.g., Homework Assigned); (b) *Model*, 12 items that assess model-specific content (e.g., Relaxation, Exposure); and (c) *Delivery*, six items that measure how model items are delivered (e.g., Modeling, Rehearsal). Coders watch entire sessions and rate each item on a 7-point Extensiveness scale (Hogue, Liddle, & Rowe, 1996): 1 (*not at all*), 3 (*somewhat*), 5 (*considerably*), 7 (*extensively*). One Total scale score and two subscale scores are created: Skills Phase (eight items; e.g., Psychoeducation, Problem Solving) and Exposure Phase (four items; e.g., Coping Plan, Exposure). The CBAY-A items have demonstrated interrater reliability, ICC(2,2), with a mean of 0.77 (SD = 0.15; range = .48–.80) and the scale, subscale, and item scores have demonstrated evidence of construct validity, including discriminating between therapists delivering ICBT across research and practice settings from therapists delivering usual care (Southam-Gerow et al., 2016). Scale and subscale scores were produced by using the item from the scale or subscale with the highest extensiveness score for each recording. This scoring strategy was used, as opposed to producing a scale or subscale score by averaging item scores, because this approach more accurately represents the way in which ICBT is delivered (i.e., the expectation that one main ICBT intervention is

TABLE 2
Cognitive-Behavioral Treatment for Anxiety in Youth Competence Scale Item Descriptions, Descriptive Data, and Interrater Reliability Results

Item Type	*Item*	*Brief Description*	*Range*	N	M	SD	*ICC (2,2)*
Standard	Within Session Focus	Therapist maintains clear focus on CBT model throughout session	6.00	682	4.65	1.36	0.76
Standard	Across Session Focus	Therapist maintains a focus on CBT model that draws across sessions	6.00	682	4.61	1.31	0.73
Standard	Structure/Phase	Therapist demonstrates competence at maintaining structure organized around typical CBT methods	6.00	682	4.56	1.37	0.78
Standard	Homework Review	Therapist skillfully reviews/assigns CBT homework	5.50	570	4.38	1.26	0.71
Standard	Homework Assignment	Therapist skillfully assigns CBT homework	6.00	522	4.29	1.36	0.75
Model	Psychoeducation-Anxiety	Therapist skillfully teaches basic information about anxiety and CBT for youth anxiety	5.00	70	3.90	0.98	0.54
Model	Emotion Education	Therapist skillfully teaches and guides the client to rehearse emotional education teaching points, with emphasis on youth anxiety.	4.50	104	4.81	1.28	0.80
Model	Fear Ladder	Therapist skillfully creates, monitors, and adjusts a hierarchy of feared stimuli to be used for exposure tasks.	5.00	82	4.45	1.05	0.64
Model	Relaxation	Therapist skillfully teaches and provides opportunities to rehearse relaxation as an anxiety management strategy.	5.50	100	4.91	1.13	0.74
Model	Cognitive-Anxiety	Therapist skillfully teaches and provides opportunity to practice CBT skills of identifying, analyzing, and/or generating alternative thoughts	5.50	63	5.02	1.24	0.71
Model	Problem Solving	Therapist skillfully teaches and provides opportunities to practice problem-solving skills for youth anxiety	4.00	30	5.40	0.94	0.67
Model	Self-Reward	Therapist skillfully teaches and provides opportunity to practice skills of (a) evaluating how well the client coped with anxiety and (b) rewarding self for coping effort	4.00	48	5.16	0.98	0.80
Model	Coping Plan	Therapist skillfully teaches and provides opportunity to practice a multistep coping plan designed to assist client to approach and cope with anxiety-provoking situations	5.00	304	4.75	1.05	0.66
Model	Exposure Preparation	Therapist skillfully and collaboratively prepares client for an exposure task	6.00	229	4.90	1.09	0.69
Model	Exposure	Therapist skillfully and collaboratively encourages client to engage in an exposure	5.50	156	4.92	1.09	0.66
Model	Exposure Debrief	Therapist skillfully and collaboratively debriefs a client after exposure	5.00	194	4.67	1.11	0.52
Model	Maintenance	Therapist skillfully and collaboratively prepares for end of treatment through (a) reviewing progress, (b) identifying and planning for future challenges, and (c) rehearsing ways to deal with the future challenges	1.50	9	4.67	0.61	0.37
Delivery	Didactic Teaching	Therapist skillfully teaches CBT for youth anxiety material in didactic and verbal manner	6.00	381	4.54	1.14	0.68
Delivery	Collaborative Teaching	Therapist skillfully teaches CBT for youth anxiety material in collaborative manner	5.50	412	4.86	1.23	0.69
Delivery	Modeling	Therapist skillfully teaches specific CBT for youth anxiety skills using observational learning methods	5.00	218	5.13	0.83	0.49
Delivery	Rehearsal	Therapist skillfully and collaboratively encourages the client to actively practice a CBT for youth anxiety skill	5.50	430	5.04	1.08	0.68
Delivery	Coaching	Therapist skillfully directs and provides feedback to a client as he/she is practicing a CBT for youth anxiety skill	3.00	53	4.86	0.68	0.55
Delivery	Self-Disclosure	The therapist skillfully shares personal information to enhance CBT for youth anxiety	5.00	137	4.82	1.00	0.63
Global	Skillfulness	Overall level of therapist skillfulness?	6.00	698	4.78	1.43	0.78
Global	Responsiveness	Overall level of therapist responsiveness?	6.00	698	4.65	1.52	0.78

Note: ICC = intraclass correlation coefficient; CBT = cognitive-behavioral treatment.

delivered per session; see Southam-Gerow et al., 2016, for more details). For the present sample, the average item interrater reliability, ICC(2,2), was 0.75 (*SD* = 0.16).

Therapy Process Observational Coding System for Child Psychotherapy—Revised Strategies scale (TPOCS-RS; McLeod, Smith, Southam-Gerow, Weisz, & Kendall, 2015) is a 42-item instrument designed to assess therapist delivery of interventions across five theory-based subscales: Cognitive (four items; e.g., Cognitive Distortions), Behavioral (nine items; e.g., Operant Interventions), Psychodynamic (five items; e.g., Interpretation), Family (seven items; e.g., Parenting Intervention), and Client-Centered (four items; e.g., Positive Regard). There are 13 additional items (e.g., Homework, Play Therapy) that represent interventions that play a meaningful role in treatment but are not associated with a specific theory-based subscale. Coders rate the extent

TABLE 3
Correlations Between CBAY-C Item, CBAY-C Global Item, TPOCS-A, and CIRS Scores

Item Type	*CBAY-C Item*	*CBAY-A Item*[a]	*Skill.*	*Respon.*	*TPOCS-A*	*CIRS*
Standard	Within Session	—	.88**	.88**	.26**	.18**
	Across Session	—	.88**	.89**	.27**	.19**
	Structure/Phase	—	.87**	.88**	.25**	.18**
	Homework Review	.50**	.71**	.73**	.21**	.14**
	Homework Assigned	.58**	.72**	.74**	.21**	.11*
Model	Psychoeducation Anxiety	.31**	.61**	.69**	.11	.10
	Emotion Education	.72**	.90**	.91**	.35**	.23*
	Fear Ladder	.48**	.68**	.71**	.25	.28*
	Relaxation	.63**	.59**	.68**	.22**	.13
	Cognitive Anxiety	.62**	.90**	.89**	.28*	.19
	Problem Solving	.56**	.88**	.86**	.16	.22
	Self-Reward	.38**	.91**	.83**	.15	.21
	Coping Plan	.37**	.80**	.83**	.21**	.22**
	Exposure: Prep	.53**	.80**	.83**	.18**	.18**
	Exposure	.43**	.81**	.81**	.18**	.08
	Exposure: Debrief	.53**	.64**	.72**	.17*	.14
	Maintenance	.17	.77*	.81**	.71**	.74*
Delivery	Didactic Teaching	.42**	.83**	.85**	.21**	.16**
	Collaborative Teaching	.49**	.83**	.87**	.28**	.23**
	Modeling	.31**	.71**	.72**	.10	.14*
	Rehearsal	.56**	.86**	.85**	.18**	.17**
	Coaching	–.06	.69**	.74**	–.002	.02
	Self-Disclosure	.06	.75**	.75**	.03	.11

Note: CBAY-C = Cognitive-Behavioral Treatment for Anxiety in Youth Competence Scale; TPOCS-A = The Therapy Process Observational Coding System for Child Psychotherapy–Alliance scale; CIRS = Child Involvement Rating Scale; CBAY-A = Cognitive-Behavioral Treatment Adherence Scale for Youth Anxiety; Skill. = Skillfulness; Respon. = Responsiveness.

[a]CBAY-A items designed to parallel the corresponding CBAY-C item and thus assess the same therapeutic intervention or delivery approach.

*$p < .05$. **$p < .01$.

to which the therapist engages in each item during an entire session on a 7-point Extensiveness scale (Hogue et al., 1996): 1 (*not at all*), 3 (*somewhat*), 5 (*considerably*), 7 (*extensively*). TPOCS-RS item scores have demonstrated acceptable interrater reliability in past studies, whereas the subscale scores have demonstrated validity evidence across different forms of treatment (e.g., McLeod & Weisz, 2010; Weisz et al., 2009). The current study used the Psychodynamic, Family, and Client-Centered subscales. We also used the Coping Cat subscale comprised of items from the Cognitive and Behavioral subscales that match content in the Coping Cat program (i.e., Relaxation, Cognitive Education, Cognitive Distortion, Coping Skills, Operant, Respondent; Southam-Gerow et al., 2010). In this study, interrater reliability, ICC (2,2), for item scores averaged .74 (*SD* = .19).

Therapy Process Observational Coding System for Child Psychotherapy-Alliance scale (TPOCS-A; McLeod & Weisz, 2005) is a nine-item scale that assesses the quality of the client–therapist alliance in youth treatment. The TPOCS-A consists of six items that assess affective elements of the client–therapist relationship and three items that assess client participation in therapeutic activities. Coders observe entire sessions and rate each item on a 6-point scale ranging from 0 (*not at all*) to 5 (*a great deal*). The TPOCS-A has demonstrated item interrater reliability ranging from .48 to .80 (*M* ICC = .67) and internal consistency ranging from .91 to .95 (*M* α = .92). Scores on the TPOCS-A scale have demonstrated evidence of convergent validity with self-report alliance instruments ranging from .48 to .53 (Fjermestad et al., 2012; Liber et al., 2010) and predictive validity with clinical outcomes (Liber et al., 2010; McLeod & Weisz, 2005). Interrater reliability, ICC (2,2), for the TPOCS-A in the present sample was .82; internal consistency was .81.

Child Involvement Rating Scale (CIRS; Chu & Kendall, 2004) is a six-item instrument that assesses aspects of positive and negative client involvement in therapeutic activities. Coders view entire sessions and then rate items on a 6-point scale ranging from 0 (*not at all*) to 5 (*a great deal*). The CIRS has demonstrated interrater reliability (ICC) from .61 to .90 and internal consistency of .73 (Chu & Kendall, 2004, 2009). Interrater reliability, ICC(2,2), for the CIRS in this study was .78; internal consistency was .85.

Coding and Session Sampling Procedures

Coders

Four doctoral students in clinical psychology composed the CBAY-C coding team (25.0% male; *M* age = 28.00, *SD* = 2.71; 50.0% Caucasian, 50.0% Latina). Each CBAY-C coder had training and clinical experience delivering CBT for anxiety. Two doctoral students (M_{age} = 27.00 years, *SD* = 4.23) in clinical psychology (50.0% male; 50.0% Caucasian, 50.0% Latina) composed the team that rated the CBAY-A and CIRS. Finally, two doctoral students (M_{age} = 24.00 years, *SD* = 1.41) in clinical psychology (100.0% female; 50.0% Caucasian, 50.0% Asian American) composed the coding team that rated the TPOCS-RS and TPOCS-A.

Coder Training

Training progressed through the same steps for each instrument. First, coders received didactic instruction and discussion of the scoring manuals, reviewed sessions with the trainers, and engaged in exercises designed to expand understanding of each item. Second, coders engaged in coding, and results were discussed in weekly meetings.

Last, coder independently coded 32 recordings and reliability was assessed against master codes produced by the first and second authors. To be certified for independent coding, each coder had to demonstrate "good" reliability on each item (ICC > .59; Cicchetti, 1994).

Assignment and Coding of Sessions

All sessions for each case were coded except the first and last session, as these sessions may contain intake or termination content. Sessions were not rated if (a) shorter than 15 min, (b) less than 15 min was audible, (c) less than 75.0% of the dialogue was in English, or (d) the recording was missing or damaged. Coding order was determined by random assignment. Each session was double-coded. Coders were naïve to study hypotheses and differences between data sources. Of the 1,098 sessions held, 744 (67.7%) were rated (65.5% ICBT; n = 532; 74.1% YAS-ICBT, n = 212). There was no significant difference between groups in the overall percentage of sessions coded, $t(66) = 1.85$, $p = .069$, or in the percentage of sessions coded from the first and second half of treatment (first half = 67.6%; second half = 67.9%), $t(67) = 0.07$, $p = .95$.

Data Analyses

Interrater Reliability

The interrater reliability was calculated for each CBAY-C item using the ICC (Shrout & Fleiss, 1979). Following Cicchetti (1994), ICCs below .40 reflect "poor" agreement, ICCs from .40 to .59 reflect "fair" agreement, ICCs from .60 to .74 reflect "good" agreement, and ICCs .75 and higher reflect "excellent" agreement. The reliability coefficients represent the model ICC(2,2) based on a two-way random effects model, which provides an estimate of the ratio of the true score variance (i.e., competence variance) to total variance (i.e., sum of competence variance and error variance). Thus, these correlations provide a reliability estimate of the mean scores of all coders considered as a whole and allow for generalizability to other samples.

Construct Validity of CBAY-C Item Scores

We evaluated whether the CBAY-C item scores demonstrated evidence of construct validity. These analyses focused on the magnitude of the correlations between scores on the CBAY-C items and corresponding adherence (CBAY-A) items, corresponding cognitive and behavioral therapeutic intervention items (TPOCS-RS), an alliance instrument (TPOCS-A), and a client involvement instrument (CIRS). We also report correlations between each discrete CBAY-C item and the two CBAY-C global items, as well as correlations between the CBAY-A and corresponding cognitive and behavioral intervention items (TPOCS-RS). The correlations were interpreted following Rosenthal and Rosnow's (1984) guidelines: r is "small" if 0.10–0.23, "medium" if 0.24–0.36, and "large" if greater than 0.36. For follow-up contrasts, standardized contrasts were calculated from the difference in the mean correlation coefficients (Hedges, 1994). The significance of each contrast was determined by first dividing the contrast value by the square root of the pooled variance, which produces a z score.

Construct Validity of CBAY-C Scale and Subscale Scores

Consistent with the CBAY-A, we created CBAY-C Total scale and subscale scores (Skills Phase, Exposure Phase) from the Model items. The Skills Phase and Exposure Phase subscales represent common phases in ICBT programs for youth anxiety, including the Coping Cat program (Kendall & Hedtke, 2006). Eight items were included on the Skills Phase subscale: Psychoeducation, Emotion Education, Fear Ladder, Relaxation, Cognitive-Anxiety, Problem Solving, Self-Reward, and Coping Plan. Four items were on the Exposure Phase subscale: Coping Plan, Exposure Prep, Exposure, and Exposure Debrief. Coping Plan was included in both, as it is prescribed in both phases. The Maintenance item was not included, because it was observed in only nine of 744 sessions. Scale and subscale scores were generated by taking the item with the highest competence score for each recording. We assessed the magnitude of the correlations between scores on the CBAY-C Total scale, two CBAY-C subscales (Skills, Exposure), two global CBAY-C items (Skillfulness, Responsiveness), and scale and subscale scores on the adherence instrument (CBAY-A), an instrument designed to assess interventions (TPOCS-RS), an alliance instrument (TPOCS-A), and a client involvement (CIRS) instrument.

Variance Components Analysis

We conducted variance components analysis of the scores from the CBAY-C Total scale, CBAY-C subscales (Skill Phase, Exposure Phase), CBAY-C global items (Skillfulness, Responsiveness), CBAY-A Total scale, and CBAY-A subscales (Skill Phase, Exposure Phase) using mixed-models procedures in SAS/STAT Software 9.4 to gauge whether targets of measurement that might impact competence and adherence influenced scores on the instruments. Our goal was to compare the proportion of variance accounted for by each factor to determine if the pattern differed across competence and adherence. Variance components analysis partitions the total variance among scores into reliable sources of variance (e.g., study group, therapist, client, time in treatment). The nested design was accounted for in the ICC calculations using mixed-model procedures (see Barber et al., 2004). A separate analysis was run for each CBAY-C and CBAY-A scale, subscale, and item. Variance components were calculated using a mixed model with restricted maximum likelihood estimation for (a) Study group, (b) Therapist (nested within study group), (c) Client

(nested within study group, therapist), (d) Time (nested within client, therapist, study group), and (e) Coder. Each factor represents a possible source of variation in integrity (Barber et al., 2004). *Study group* reflects the influence of the ICBT groups (ICBT, YAS-ICBT); *therapist* represents differences across therapists; *client* reflects differences across each client; *time* reflects change in scores over treatment (measured in weeks since intake); *coder* reflects differences in coder ratings (tendency to score high or low). Variance estimates were transformed into proportions of variance based on estimates of the total variance. We also evaluated whether there were differences between ICBT and YAS-ICBT on the CBAY-C scale, subscale, and item scores. Using the same model, we computed adjusted least square means scores using SAS/STAT Software 9.4. Because our primary interest was to conduct a group comparison, we examined the overall F test for group with a Bonferroni adjusted alpha of .01.

RESULTS

Interrater Reliability

ICCs ranged from 0.37 to 0.80 (M = 0.67, SD = 0.11; see Table 2). The ICCs for seven of the 25 items fell within the "excellent" range, 13 items fell within the "good" range, four fell within the "fair" range, and one item (Maintenance) fell within the "poor" range (Cicchetti, 1994). Twenty-three items displayed nearly the full range of scores, whereas two items used less than half of the 7-point competence scale: Maintenance (range = 1.5) and Coaching (range = 3).

Construct Validity of the CBAY-C Items

As seen in Tables 3 and 4, the magnitude of the interitem correlations among the CBAY-C (competence), CBAY-A (adherence), TPOCS-RS (cognitive and behavioral interventions), TPOCS-A (alliance), and CIRS (client involvement) items ranged from small to large. The correlations between the CBAY-C discrete and global competence items were "large" in magnitude with a range from .59 to .91 (M = .78, SD = .10) for Skillfulness and .68 to .91 (M = .80, SD = .07) for Responsiveness. The correlations between the CBAY-A and TPOCS-RS items ranged from .25 to .88 (M = .60, SD = .19), with all but three correlations considered large (r > .36; Rosenthal & Rosnow, 1984). Interitem correlations between scores on the CBAY-C items and the corresponding CBAY-A items ranged from –.06 to .72 (M = .43, SD = .20), with all but six coefficients in the "large" range. Of note, three of the correlations fell below .23: Maintenance (r = .17, *ns*), Self-Disclosure (r = .06, *ns*), and Coaching (r = –.06, *ns*). The interitem correlations between scores on the CBAY-C items and corresponding TPOCS-RS items (M = .34, SD = .17; range = .08–.65) were all lower than the CBAY-A correlations, with more than half of the correlations falling in the "small" to "medium" range. The interitem correlations between the CBAY-C items and the TPOCS-A ranged from .00 to .71 (M = .21, SD = .14), with 15 of the 21 correlations considered "small" (r < .23; Rosenthal & Rosnow, 1984). For the CIRS, the interitem correlations ranged from .02 to .28 (M = .19, SD = .13), with all but one correlation considered "small." The mean of the correlations between the CBAY-A and corresponding TPOCS-RS items (M = .60, SD = .19), designed to assess treatment adherence, were stronger than the mean of the

TABLE 4
Correlations Between Corresponding TPOCS-RS, CBAY-C, and CBAY-A Item Scores

Item Type	*CBAY-C Item*	*TPOCS-RS Item*	*CBAY-C*	*CBAY-A*
Standard	Homework Review	Homework	.42**	.68**
	Homework Assigned	Homework	.47**	.64**
Model	Psychoeducation Anxiety	Psychoeducation	.09	.27**
	Emotion Education	Cognitive Education	.51**	.52**
	Relaxation	Relaxation	.65**	.88**
	Cognitive Anxiety	Cognitive Distortion	.51**	.53**
	Problem Solving	Coping Skills	.20	.36**
	Self-Reward	Operant	.48**	.59**
	Coping Plan	Coping Skills	.32**	.71**
	Exposure: Prep	Respondent	.40**	.82**
	Exposure	Respondent	.30**	.70**
	Exposure: Debrief	Respondent	.25**	.74**
Delivery	Modeling	Modeling	.13	.65**
	Rehearsal	Rehearsal	.35**	.81**
	Coaching	Coaching	.22	.25**
	Self-Disclosure	Self-Disclosure	.08	.50**

Note: TPOCS-RS = Therapy Process Observational Coding System for Child Psychotherapy–Revised Strategies scale; CBAY-C = Cognitive-Behavioral Treatment for Anxiety in Youth Competence Scale; CBAY-A = Cognitive-Behavioral Treatment Adherence Scale for Youth Anxiety.

*p < .05. **p < .01.

correlations between the CBAY-C and CBAY-A items ($z = 7.37$, $p < .001$). The mean of the correlations between the CBAY-C and CBAY-A items were stronger than the mean of the correlations between the CBAY-C and the TPOCS-A ($z = 16.93$, $p < .001$) and the CIRS ($z = 21.51$, $p < .001$). The correlations were thus uniformly strong for the instruments that assess the same therapeutic content (CBAY-A, TPOCS-RS) and lowest for the instruments that assess distinct aspects of the treatment process (CBAY-C, TPOCS-A, CIRS). However, the pattern of correlations did not support the construct validity of the Maintenance, Coaching, or Self-Disclosure items. Altogether, these findings indicate that the CBAY-C item scores overlapped to a moderate degree with adherence instruments and to a lesser degree with instruments assessing alliance and involvement.

Construct Validity of the CBAY-C Scale, Subscales, and Global Items

As seen in Table 5, the magnitude of the correlations among the scores on five observational instruments (CBAY-C, CBAY-A, TPOCS-RS, TPOCS-A, CIRS) ranged from small to large. The strongest correlations were observed between scores on the two CBAY-C global items, followed by the CBAY-C scale and subscales. In fact, the magnitude of the correlations suggests that these scores may be redundant ($r > .70$; Kline, 1979). The correlations between scores on the CBAY-C subscales and the CBAY-A subscales ranged from .22 to .65 ($M = .46$, $SD = .17$) and the TPOCS-RS Coping Cat subscale ranged from .44 to .51 ($M = .48$, $SD = .04$). The CBAY-C global item correlations were of similar magnitude. The correlations between scores on the CBAY-C global items ranged from .42 to .71 ($M = .56$, $SD = .12$) with the CBAY-A scale/subscales and ranged from .61 to .62 ($M = .62$, $SD = .01$) with the TPOCS-RS Coping Cat subscale. The remaining correlations fell in the small to medium range, save for one correlation between the CBAY-C Responsiveness item and the TPOCS-A ($r = .38$). The mean of the correlations between the CBAY-A subscales and the TPOCS-RS Coping Cat subscale, designed to assess treatment adherence, was higher than the mean of the correlations between the CBAY-C and CBAY-A subscales ($z = 2.71$, $p < .01$). The mean of the correlations between the CBAY-C and CBAY-A subscales was higher than the mean of the correlations between the CBAY-C subscales and the TPOCS-A ($z = 5.43$, $p < .001$) and the CIRS ($z = 8.67$, $p < .001$). Together, these findings support the construct validity of the CBAY-C scores, though the findings do suggest that the global items may be redundant ($r > .70$; Kline, 1979).

Variance Components Analysis

As seen in Table 6, study group accounts for a substantial proportion of the variance in CBAY-C scale, subscale, and global item scores, suggesting that the competence scores varied across the ICBT groups. Time in treatment accounted for a slightly smaller proportion of the variance, indicating that the competence scores may vary over treatment. Of interest, the opposite pattern was seen in the CBAY-A scale and subscale scores such that the highest proportion of variance was accounted for by time in treatment with study group accounting for a smaller proportion. As might be expected, therapist did account for a substantial

TABLE 5
Correlations Between the CBAY-C Scale, CBAY-C Subscale, CBAY-C Global Items, CBAY-A Scale, CBAY-A Subscale, TPOCS-RS Subscales, the TPOCS-A, and CIRS Scores

	2.	*3.*	*4.*	*5.*	*6.*	*7.*	*8.*	*9.*	*10.*	*11.*	*12.*	*13.*	*14.*
1. CBAY-C Skills	.86**	.97**	.87**	.88**	.57**	.22**	.65**	.49**	.39**	−.22**	−.29**	.27**	.21**
2. CBAY-C Exposure		.95**	.83**	.84**	.22**	.55**	.55**	.44**	.31**	−.22**	−.17**	.21**	.20**
3. CBAY-C Total Model			.88**	.89**	.43**	.33**	.65**	.51**	.38**	−.23**	−.23**	.25**	.21**
4. CBAY-C Skill. Item				.95**	.52**	.45**	.71**	.62**	.38**	−.27**	−.35**	.34**	.20**
5. CBAY-C Respon. Item					.53**	.42**	.70**	.61**	.38**	−.29**	−.37**	.38**	.24**
6. CBAY-A Skills						.08**	.81**	.56**	.33**	−.29**	−.29**	.30**	.20**
7. CBAY-A Exposure							.45**	.48**	.24**	−.16**	−.16**	.20**	.13**
8. CBAY-A Total Model								.77**	.40**	−.43**	−.35**	.38**	.23**
9. TPOCS-RS CC									.40**	−.45**	−.28**	.35**	.18**
10. TPOCS-RS Client										−.25**	−.16**	.39**	.27**
11. TPOCS-RS Family											.14**	−.18**	−.18**
12. TPOCS-RS Dynamic												−.46**	−.31**
13. TPOCS-A													.77**
14. CIRS													

Note: CBAY-C = Cognitive-Behavioral Treatment for Anxiety in Youth Competence Scale; CBAY-A = Cognitive-Behavioral Treatment Adherence Scale for Youth Anxiety; TPOCS-RS = Therapy Process Observational Coding System for Child Psychotherapy–Revised Strategies Scale; TPOCS-A = Therapy Process Observational Coding System for Child Psychotherapy–Alliance scale; CIRS = Child Involvement Rating Scale; Skill. = Skillfulness; Respon. = Responsiveness; CC = Coping Cat; Client = client-centered; Dynamic = psychodynamic.

**$p < .01$.

TABLE 6
Variance Components for CBAY-C Scale, CBAY-C Subscale, CBAY-C Global Item, CBAY-A Scale, and CBAY-A Subscale Scores

	Variance Components					
Scale or Subscale	*Group*	*Therapist*	*Client*	*Time*	*Coder*	*Residual*
CBAY-C Skills Subscale	.34	.14	< .01	.17	.09	.26
CBAY-C Exposure Subscale	.43	.11	< .01	.15	.09	.22
CBAY-C Total Scale	.42	.13	< .01	.16	.05	.23
CBAY-C Skillfulness	.54	.11	.02	.13	.02	.18
CBAY-C Responsiveness	.48	.12	.02	.14	.02	.22
CBAY-A Skills Subscale	.07	.12	< .01	.46	< .01	.34
CBAY-A Exposure Subscale	.22	.02	.01	.55	.01	.18
CBAY-A Total Scale	.25	.12	.02	.39	.03	.19

Note: CBAY-C = Cognitive-Behavioral Treatment for Anxiety in Youth Competence Scale; CBAY-A = Cognitive-Behavioral Treatment Adherence Scale for Youth Anxiety.

TABLE 7
Least Square Means of CBAY-C Scale, Subscale, and Global Item Scores Across Groups

CBAY-C Scale, Subscale, and Items	M *ICBT*	M *YAS-ICBT*	F	p
Skills Subscale	5.14	3.51	5.41	.020
Exposure Subscale	5.14	3.41	7.10	.008*
Total Scale	5.33	3.55	5.73	.017
Skillfulness Item	5.26	2.69	5.04	.024
Responsiveness Item	5.13	2.61	8.56	.004*

Note: Pairwise comparisons were conducted with a Bonferonni adjusted alpha. CBAY-C = Cognitive-Behavioral Treatment of Anxiety in Youth Competence Scale; ICBT = individual cognitive-behavioral treatment delivered in Kendall et al. (2008) study; YAS-ICBT = individual cognitive-behavioral delivered in Southam-Gerow et al. (2010) study.

*$p < .003$.

proportion of variance in the scores on the CBAY-C and CBAY-A, suggesting that some therapists had higher competence and adherence scores than other therapists. Because study group accounted for a substantial proportion of the variance in CBAY-C scores we conducted group comparisons. As seen in Table 7, the ICBT group had higher scores than the YAS-ICBT group on each scale, subscale, and global item. However, after Bonferroni correction, only the Exposure subscale and Responsiveness item were significantly higher.

DISCUSSION

This study reports on the development of a therapist competence instrument for ICBT for youth anxiety, the CBAY-C, and presents initial reliability and representative validity evidence for the scale, subscale, and item scores. A set of trained coders produced encouraging interrater reliability on the CBAY-C item scores. Findings also demonstrated that the scale, subscale, and item scores were distinct from independent ratings of adherence, cognitive-behavioral interventions, the alliance, and client involvement, providing preliminary validity evidence.

Our results indicate that our observational instrument that gauges therapist competence in ICBT for youth anxiety can be coded reliably by independent coders rating sessions delivered in research and community settings. The interrater reliability data reported herein compares favorably to that of other observational competence instruments (e.g., Bjaastad et al., 2016; Hogue et al., 2008a). Our findings also provide initial evidence supporting the representative validity of the CBAY-C scale, subscale, and item scores. For example, CBAY-C scale, subscale, and item scores demonstrated medium to strong correlations with the same item, scale, and subscales measured by the CBAY-A. The strength of these correlations was consistent with previous research reporting on the correspondence between competence and adherence (Carroll et al., 2000; Hogue et al., 2008a), though lower than some previous estimates (e.g., Barber et al., 2003). Our findings also provide initial discriminant validity evidence as the CBAY-C scale, subscale, and item scores demonstrated low to medium correlations with observational instruments of the alliance and client involvement. Again, these findings are consistent with previous research reporting on the correlation between competence and the alliance (Carroll et al., 2000; Hogue et al., 2008a).

In developing the CBAY-C we addressed an important measurement question: Is there overlap between adherence and competence? To address this question, we investigated the extent to which the CBAY-C overlapped with the CBAY-A, an observational adherence instrument of ICBT for youth anxiety designed to parallel the content of the CBAY-C. Our findings suggest that the CBAY-C and CBAY-A scores are distinct. First, only one correlation between the CBAY-C and corresponding CBAY-A item exceeded .70 (i.e., Emotion Education; $r = .72$) suggesting the items were not redundant (Kline, 1979). Second, the mean level of correlation coefficients ($r = .53$) was lower

than found in a recent study of similar instruments for CBT for youth anxiety (r = .79; Bjaastad et al., 2016). Third, the mean of the correlations between the two adherence instruments (i.e., CBAY-A, TPOCS-RS) were more strongly related than either was with the CBAY-C. Overall, these findings suggest that the CBAY-C and CBAY-A assess distinct treatment integrity components.

Another measurement issue we sought to address was the best way to assess competence: using global or discrete competence ratings. We found that scores on the global items were strongly related to the discrete items, suggesting that these item scores may provide some unique information. However, scores on the global items were largely redundant with one another (r = .95), and the CBAY-C scale and subscale scores (rs = .83–.89). This indicates that there may be some utility in combining the global items and having coders make a single global competence rating. However, though the data suggest substantial overlap, we believe that there is conceptual value in retaining the scale and subscales for future studies as they map onto common ICBT phases. Moreover, it may be premature to drop the scale and subscales based on a single study. This study was designed to provide initial evidence for the representative validity of the CBAY-C item, scale, and subscales scores. Future studies can ascertain whether the discrete and global ratings evidence different patterns of elaborative validity evidence, particularly predictive validity with treatment outcomes.

Several factors accounted for systematic variation in the CBAY-C scores. As hypothesized, our findings suggest that scores on the CBAY-C differed across therapists. That is, some therapists deliver ICBT with higher quality than others. Time in treatment also accounted for a proportion of nonerror variance. This suggests that the CBAY-C scores may systematically change over time and is consistent with evidence that suggest scores on adherence and competence instruments vary over treatment (Barber et al., 2004; Hogue et al., 2008a). Our data, though preliminary, provide some reason to question the standard in the field to sample a single session for competence coding (Dennhag, Gibbons, Barber, Gallop, & Crits-Christoph, 2012). If competence scores systematically increase (or decrease) over treatment, such a sampling approach may be inadequate.

Coder effects were also observed, suggesting that some coders had a tendency to rate competence higher (or lower) than others. A persistent issue in the field has been the level of training needed for coding competence (Barber et al., 2007). We chose to use advanced clinical psychology graduate students with experience in CBT for anxiety. Although they obtained acceptable rates of interrater reliability, our findings indicate that there may have been some systematic differences across coders. Client effects were not observed, suggesting that competence did not vary across clients. However, as there were not many clients nested in therapists (M = 2.34, SD = 2.19), our ability to detect client (and therapist) effects was limited. Given the potential impact of client (e.g., resistance) and therapist (e.g., experience, training; Barber et al., 2007) factors on competence this is an area worthy of further study. Overall, our variance components analysis suggests that the CBAY-C scale, subscale, and item scores may be sensitive to variations across groups, therapists, time, and coder.

Study group accounted for a high proportion of nonerror variance in the CBAY-C scores. Follow-up analyses revealed that ICBT delivered in research settings had significantly higher competence scores on the Exposure subscale and the Responsiveness item. Moreover, all CBAY-C scores were higher in the ICBT condition. Why did we find this pattern? One possibility is that therapists in the ICBT group were trained by the treatment developer and delivered the treatment in a research setting. As a result, variables like therapist selection and content expertise may account for the higher scores. Therapists in the research setting were more likely to complete the exposure tasks, which could influence competence ratings. It is plausible that therapists in community settings trained to provide ICBT will demonstrate lower competence with their first cases than would therapists seeing clients in a research clinic overseen by the treatment developer. Also, there is evidence that the clients seen in research clinics differ in ways that could potentially influence competence. For example, clients in research settings tend to be more ethnically homogeneous and have lower rates of externalizing psychopathology, compared to youth seen in community clinics (see Southam-Gerow, Chorpita, Miller, & Gleacher, 2008). These differences may drive some of the variance in competence scores observed here.

The variance components analysis provided evidence that the CBAY-C and CBAY-A scale, subscale, and items tap into distinct integrity components. Different factors accounted for the highest proportion of nonerror variance in the CBAY-A and CBAY-C scores, suggesting that the two instruments are picking up different patterns of variation. This suggests that an important direction for future research is to identify whether distinct factors do account for different patterns of variation in adherence and competence scores. Such research may help researchers pinpoint ways to maintain integrity in various settings (Boswell et al., 2013).

Although most results support the psychometric profile of the CBAY-C scores, three items did not perform as expected. Specifically, our analyses did not support the validity of the Self-Disclosure and Coaching item scores. Future work with the CBAY-C should not use the Self-Disclosure or Coaching items unless they are refined and only if there is a substantial conceptual reason to retain them. Neither item demonstrated the full range of scores, suggesting that the scoring guidelines will need to be revised to improve the range and help improve score validity. In addition, the Maintenance item did not have

acceptable interrater reliability (ICC = .37) or score validity. This is likely due to the low number of observations (n = 9), which may be due to the fact this intervention is not emphasized in the Coping Cat manual. Further, our choice not to code the last session may have influenced our ability to observe the intervention. The strength of the Maintenance item should be evaluated in future research with ICBT programs that include Maintenance (e.g., Modular Approach for Treating Childhood Problems manual; Chorpita & Weisz, 2009).

The findings have implications for future research. The CBAY-C is a strong candidate for implementation research. As part of a quality control system, the CBAY-C could support efforts to evaluate the impact of training and supervision on therapist competence. Given differences noted across sites, multisite studies may consider using the CBAY-C to evaluate the extent to which therapist competence is comparable across sites. Finally, to our knowledge the CBAY-C is the first competence instrument comprising practice elements found across ICBT programs, as opposed to interventions found in specific treatment manuals. This suggests that it may be possible for future work to develop treatment integrity instruments that can apply across multiple treatment programs (McLeod et al., 2013), which could help enhance the practicality of integrity measurement.

Although the study contains a number of strengths, including a large sample of recordings from two studies all double-coded by trained coders, there are limitations to consider. First, until recently there was not another competence instrument for ICBT for youth anxiety (see Bjaastad et al., 2016). As a result, our validity analyses lack a true comparator instrument for convergent validity. For this reason, it will be important for future research to assess the convergent validity of the CBAY-C scores. Second, our study did not include therapists with extensive expertise in ICBT for youth anxiety, so the performance of the scale, subscale, and item scores with experienced therapists is unknown. Third, ICBT for youth anxiety often includes exposure interventions, some of which occur outside of treatment rooms. Because we relied on recordings, it is possible that we missed some of the exposure delivery, making our examination of this item somewhat incomplete. Finally, although our sample included a diverse set of youth and therapists, no sample can completely represent how treatment is delivered in research or clinical settings, so these findings should be replicated.

Despite these limitations, the present study provides initial data supporting the score reliability and validity on an observational instrument of competence for ICBT for youth anxiety, the CBAY-C. The CBAY-C items can be coded reliably across four item categories (Standard, Model, Delivery, Global). Moreover, preliminary evidence supports the representative validity of the items scores (Foster & Cone, 1995), suggesting that they are distinct from ratings of treatment adherence. An important next step will be to investigate the convergent validity as well as the elaborative validity of the CBAY-C scores by investigating relations with treatment outcomes in ICBT for youth anxiety.

FUNDING

Preparation of this article was supported in part by a grant from the National Institute of Mental Health Grant (RO1 MH086529; McLeod & Southam-Gerow).

REFERENCES

Barber, J. P., Foltz, C., Crits-Christoph, P., & Chittams, J. (2004). Therapists' adherence and competence and treatment discrimination in the NIDA collaborative cocaine treatment study. *Journal of Clinical Psychology*, *60*(1), 29–41. doi:10.1002/(ISSN)1097-4679

Barber, J. P., Liese, B. S., & Abrams, M. J. (2003). Development of the cognitive therapy adherence and competence scale. *Psychotherapy Research*, *13*(2), 205–221. doi:10.1093/ptr/kpg019

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*(5), 493–500. doi:10.1037/0735-7028.38.5.493

Bjaastad, J. F., Haugland, B. S. M., Fjermestad, K. W., Torsheim, T., Havik, O. E., Heiervang, E. R., & Öst, L.-G. (2016). Competence and adherence scale for cognitive behavioral therapy (CAS-CBT) for anxiety disorders in youth: Psychometric properties. *Psychological Assessment*, *28*(8), 908–916. doi:10.1037/pas0000230

Boswell, J. F., Gallagher, M. W., Sauer-Zavala, S. E., Bullis, J., Gorman, J. M., Shear, M. K., … Barlow, D. H. (2013). Patient characteristics and variability in adherence and competence in cognitive-behavioral therapy for panic disorder. *Journal of Consulting and Clinical Psychology*, *81*(3), 443–454. doi:10.1037/a0031437

Carroll, K. M., Nich, C., Sifry, R. L., Nuro, K. F., Frankforter, T. L., Ball, S. A., … Rounsaville, B. J. (2000). A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence*, *57*(3), 225–238. doi:10.1016/S0376-8716(99)00049-6

Castonguay, L. G. (1993). "Common factors" and "nonspecific variables": Clarification of the two concepts and recommendations for research. *Journal of Psychotherapy Integration*, *3*(3), 267–286. doi:10.1037/h0101171

Chorpita, B. F., & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology*, *77*(3), 566–579. doi:10.1037/a0014565

Chorpita, B. F., & Weisz, J. R. (2009). *Modular approach to therapy for children with anxiety, depression, traumatic stress, and conduct problems (MATCH-ADTC)*. Satellite Beach, FL: PracticeWise.

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*(5), 821–829. doi:10.1037/0022-006X.72.5.821

Chu, B. C., & Kendall, P. C. (2009). Therapist responsiveness to child engagement: Flexibility within manual-based CBT for anxious youth. *Journal of Clinical Psychology*, *65*(7), 736–754. doi:10.1002/jclp.v65:7

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. doi:10.1037/1040-3590.6.4.284

Dennhag, I., Gibbons, M. B. C., Barber, J. P., Gallop, R., & Crits-Christoph, P. (2012). How many treatment sessions and patients are needed to create a stable score of adherence and competence in the treatment of cocaine dependence? *Psychotherapy Research*, *22*(4), 475–488. doi:10.1080/10503307.2012.674790

Fjermestad, K. W., McLeod, B. D., Heiervang, E. R., Havik, O. E., Öst, L.-G., & Haugland, B. S. M. (2012). Factor structure and validity of the therapy process observational coding system for child psychotherapy – alliance scale. *Journal of Clinical Child & Adolescent Psychology*, *41*, 246–254. doi:10.1080/15374416.2012.651999

Forgatch, M. S., Patterson, G. R., & Degarmo, D. S. (2005). Evaluating fidelity: Predictive validity for a measure of competent adherence to the oregon model of parent management training. *Behavior Therapy*, *36*(1), 3–13. doi:10.1016/S0005-7894(05)80049-8

Foster, S. L., & Cone, J. D. (1995). Validity issues in clinical assessment. *Psychological Assessment*, *7*(3), 248–260. doi:10.1037/1040-3590.7.3.248

Garland, A. F., & Schoenwald, S. K. (2013). Use of effective and efficient quality control methods to implement psychosocial interventions. *Clinical Psychology: Science and Practice*, *20*(1), 33–43.

Goense, P., Boendermaker, L., Van Yperen, T., Stams, G.-J., & Van Laar, J. (2014). Implementation of treatment integrity procedures. *Zeitschrift Für Psychologie*, *222*(1), 12–21. doi:10.1027/2151-2604/a000161

Gutermann, J., Schreiber, F., Matulis, S., Stangier, U., Rosner, R., & Steil, R. (2014). Therapeutic adherence and competence scales for developmentally adapted cognitive processing therapy for adolescents with PTSD. *European Journal of Psychotraumatology*, *1*, 1–21.

Hedges, L. V. (1994). Fixed effect models. In H. E. Cooper, & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 285−299). New York, NY: Russell Sage Foundation.

Hogue, A., Dauber, S., Chinchilla, P., Fried, A., Henderson, C., Inclan, J., … Liddle, H. A. (2008a). Assessing fidelity in individual and family therapy for adolescent substance abuse. *Journal of Substance Abuse Treatment*, *35*(2), 137–147. doi:10.1016/j.jsat.2007.09.002

Hogue, A., Henderson, C. E., Dauber, S., Barajas, P. C., Fried, A., & Liddle, H. A. (2008b). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology*, *76*(4), 544–555. doi:10.1037/0022-006X.76.4.544

Hogue, A., Liddle, H. A., & Rowe, C. (1996). Treatment adherence process research in family therapy: A rationale and some practical guidelines. *Psychotherapy: Theory, Research, Practice, Training*, *33*(2), 332–345. doi:10.1037/0033-3204.33.2.332

Kendall, P. C. (1994). Treating anxiety disorders in children: Results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology*, *62*, 100–110. doi:10.1037/0022-006X.62.1.100

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 disordered youth: A randomized clinical trial evaluating child and family modalities. *Journal of Consulting and Clinical Psychology*, *76*(2), 282–297. doi:10.1037/0022-006X.76.2.282

Kline, P. (1979). *Psychometrics and psychology*. London, UK: Academic Press.

Liber, J. M., McLeod, B. D., Van Widenfelt, B. M., Goedhart, A. W., Van Der Leeden, A. J. M., Utens, E. M. W. J., & Treffers, P. D. A. (2010). Examining the relation between the therapeutic alliance, treatment adherence, and outcome of cognitive behavioral therapy for children with anxiety disorders. *Behavior Therapy*, *41*(2), 172–186. doi:10.1016/j.beth.2009.02.003

McLeod, B. D., Smith, M. M., Southam-Gerow, M. A., Weisz, J. R., & Kendall, P. C. (2015). Measuring treatment differentiation for implementation research: The therapy process observational coding system for child psychotherapy revised strategies scale. *Psychological Assessment*, *27*(1), 314–325. doi:10.1037/pas0000037

McLeod, B. D., Southam-Gerow, M. A., Bair, C. E., Rodriguez, A., & Smith, M. M. (2013). Making a case for treatment integrity as a psychological treatment quality indicator. *Clinical Psychology: Science and Practice*, *20*(1), 14–32.

McLeod, B. D., & Weisz, J. R. (2005). The Therapy Process Observational Coding System-Alliance scale: Measure characteristics and prediction of outcome in usual clinical practice. *Journal of Consulting and Clinical Psychology*, *73*(2), 323–333. doi:10.1037/0022-006X.73.2.323

McLeod, B. D., & Weisz, J. R. (2010). The Therapy Process Observational Coding System for Child Psychotherapy-Strategies scale. *Journal of Clinical Child and Adolescent Psychology*, *39*(3), 436–443. doi:10.1080/15374411003691750

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice*, *12*(4), 365–383.

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, *75*, 829–841. doi:10.1037/0022-006X.75.6.829

Rodolfa, E., Bent, R., Eisman, E., Nelson, P., Rehm, L., & Ritchie, P. (2005). A cube model for competency development: Implications for psychology educators and regulators. *Professional Psychology: Research and Practice*, *36*(4), 347–354. doi:10.1037/0735-7028.36.4.347

Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York, NY: McGraw-Hill.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. doi:10.1037/0033-2909.86.2.420

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*(3), 168–180. doi:10.1007/s10488-007-0154-7

Southam-Gerow, M. A., McLeod, B. D., Arnold, C. A., Rodriguez, A., Cox, J. R., Reise, S. P. … Kendall, P. C. (2016). Initial development of a treatment adherence measure for cognitive-behavioral therapy for child anxiety. *Psychological Assessment*, *28*(1), 70–80. doi: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 & Adolescent Psychiatry*, *49*(10), 1043–1052. doi:10.1016/j.jaac.2010.06.009

Stallard, P., Myles, P., & Branson, A. (2014). The Cognitive Behaviour Therapy Scale for Children and Young People (CBTS-CYP): Development and psychometric properties. *Behavioural and Cognitive Psychotherapy*, *42*(3), 269–282. doi:10.1017/S135246581300115X

Webb, C. A., Derubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, *78*(2), 200–211. doi:10.1037/a0018912

Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S., & Miranda, J.; Research Network on Youth Mental Health. (2012). Testing standard and modular designs for psychotherapy treating depression, anxiety, and conduct problems for youth. *Archives of General Psychiatry*, *69*(2), 274–282. doi:10.1001/archgenpsychiatry.2011.147

Weisz, J. R., Southam-Gerow, M. A., Gordis, E. B., Connor-Smith, J. K., Chu, B. C., Langer, D. A., … Weiss, B. (2009). Cognitive behavioral therapy versus usual clinical care for youth depression: An initial test of transportability to community clinics and clinicians. *Journal of Consulting and Clinical Psychology*, *77*(3), 383–396. doi:10.1037/a0013877