Received: 22 December 2016 | Revised: 7 August 2017 | Accepted: 16 August 2017

DOI: 10.1002/jclp.22537

WILEY

RESEARCH ARTICLE

The global therapist competence scale for youth psychosocial treatment: Development and initial validation

Ruth C. Brown[1] | **Michael A. Southam-Gerow**[1] | **Bryce D. McLeod**[1] | **Emily B. Wheat**[1] | **Carrie B. Tully**[1] | **Steven P. Reise**[2] | **Philip C. Kendall**[3] | **John R. Weisz**[4]

[1]Virginia Commonwealth University

[2]University of California, Los Angeles

[3]Temple University

[4]Harvard University

Correspondence
Ruth C. Brown, Virginia Commonwealth University, Virginia Institute for Psychiatric and Behavioral Genetics, 800 East Leigh St., Suite 101, Richmond, VA 23219-1534.
Email:ruth.brown@vcuhealth.org

Funding information

This research was supported by a grant from the National Institute of Mental Health Grant (RO1 MH086529; Southam-Gerow & McLeod).

Note that this study was reviewed and approved by the institutional review board (IRB) at Virginia Commonwealth University as well as the IRBs of the two institutions (i.e., University of California, Los Angeles, Temple University) at which the original data were collected.

All participants completed informed consent or assent procedures.

None of the authors have a conflict of interest

Abstract

Objective We describe the development and initial psychometric properties of the observer-rated Global Therapist Competence Scale for Youth Psychosocial Treatment (G-COMP) in the context of cognitive-behavioral treatment (CBT) for youth anxiety disorders.

Method Independent coders rated 744 sessions from a sample of 68 youth (mean age = 10.56 years) using the G-COMP and the instruments of alliance, involvement, CBT adherence, CBT competence.

Results Inter-rater reliability coefficients, ICC(2,2), were greater than .60 for the 5 G-COMP domain scores. G-COMP scores yielded small to medium correlations with instruments of alliance (*rs* = .17–.44) and youth involvement in treatment (*rs* = .08–.53), and medium to large correlations with instruments of CBT competence and adherence (*rs* = .26–.63). Therapists in the research setting were rated higher compared to newly trained therapists in community clinics.

Conclusion Preliminary reliability and validity of the G-COMP are promising, but future research is needed with non-CBT samples.

KEYWORDS

assessment, cognitive-behavioral treatment, therapist competence, treatment process, youth anxiety

Healthcare in the United States is in an era of accountability, an emphasis with direct implications for mental health services. Stakeholders invested in the outcomes of mental health services are demanding that delivery of psychosocial treatments (hereafter called *treatment*) meet certain criteria, especially related to service quality (e.g., Institute of Medicine, 2015). The emphasis on quality has arisen, in part, as a response to the increasing focus on accountability and the central role of competency in the education and practice of psychology (e.g., American Psychological

Association [APA], 2015). The goal is to improve the quality of mental health care by improving training and practice guidelines related to the delivery of treatments.

Therapist competence represents one important aspect of service quality. Efforts to define the competencies associated with the delivery of treatments have been divided into two broad categories (Barber, Sharpless, Klostermann, & McCarthy, 2007). The first, *technical competence*, focuses on competence in delivering particular therapeutic interventions found in specific treatment models or programs (also called *model-specific* and *limited domain* competence; Barber et al., 2007). Examples of instruments for measuring technical competence include ones for adult cognitive-behavioral treatment (CBT) for depression (Vallis, Shaw, & Dobson, 1986) and anxiety (Boswell et al., 2013) and family treatment and CBT for adolescent substance use (Hogue et al., 2008). The second category focuses on competencies associated with *general* therapeutic skills, called variously *common-factors* (e.g., Castonguay, 1993), *foundational* (Rodolfa et al., 2005; Spruill et al., 2004), and *global* (Barber et al., 2007) competence, that are believed to apply across treatment models.

To date, studies examining the relation between competence and treatment outcome have found mixed findings (Barber et al., 2006; Hogue et al., 2008). It is possible that the mixed findings are in part because of inadequate measurement of competence (e.g., low interrater reliability) or the use of highly trained therapists in controlled trials restricting the range of competency (Hogue et al., 2008). There are few tools specifically designed to assess these competencies in youth treatment (see Fjermestad, McLeod, Tully, & Liber, 2016).

Of the four instruments designed to assess technical therapist competence in youth treatment (Forgatch, Patterson, & Degarmo, 2005; Gutermann et al., 2014; Hogue et al., 2008; Stallard, Myles, & Branson, 2014), only three of these contain items that tap into global competence (Forgatch et al., 2005; Gutermann et al., 2014; Stallard et al., 2014). Scores on three of the observational instruments have demonstrated wide variability in inter-rater item reliability (Kappa > .69; Forgatch et al., 2005; intraclass correlation coefficients [ICCs] range from .01 to .87; Gutermann et al., 2014; Hogue et al., 2008). Studies have often found high correlations between items assessing technical and global competence (e.g., *r*s = .79–.91; Stallard et al., 2014) calling into question whether these represent distinct facets of competence or an artifact of using the same coders to rate both sets of items (i.e., halo effects). Thus, key to establishing the score validity of a global competence instrument is to demonstrate that it is distinct from technical competence.

Our study contributes to the literature by reporting on our effort to develop and distinguish an instrument that assesses a therapist's global competence. As a first step, we developed a stand-alone global competence instrument, the Global Therapist Competence Scale for Youth Psychosocial Treatment (G-COMP), and assessed its discriminant validity from other aspects of the treatment process. In this study, we report our initial psychometric findings, including a description of the development of the G-COMP and an examination of the psychometric properties of the G-COMP domain scores using two samples of youth diagnosed with a principal anxiety disorder and receiving individual CBT (ICBT), one sample treated in a clinical research setting, the other treated in community settings. This effort to establish the representative validity of a global competence instrument may help downstream efforts to develop training and practice guidelines for youth therapists.

1 | METHOD

1.1 | Instrument development

1.1.1 | Scale development

We began with the five competence domains and set out to identify empirically supported therapist behaviors, which led to the formation of each conceptual domain. Based on literature review, we identified five domains that became G-COMP scales. Table 1 provides a summary of the most prominent efforts to identify global domains, beginning with Frank (1971) and including the work of the APA's Division 29 Task Force on Empirically Supported Relationships (Norcross, 2002, 2011) and the APA's Division 12 Task Force on Empirically Based Principles of Therapeutic Change (Castonguay & Beutler, 2006). We began by sorting the individual elements from our review that lend themselves to

TABLE 1 Synthesis of theories for the global therapist competence scale for youth psychosocial treatment

G-COMP factors	Frank (1971)	Orlinsky and Howard (1987)	Grencavage & Norcross (1990)	Norcross (2002, 2011)	Castonguay and Beutler (2006)
Alliance building	• Emotionally charged, confiding relationship	• Therapeutic bond • Therapeutic openness-involvement • Therapeutic contract	• Beneficial therapist qualities • Therapeutic Environment	• Congruence/ genuineness • Repair of the alliance rupture • Positive regard • Empathy/conveys understanding • Self-disclosure • Goal consensus/ collaboration • Eliciting client feedback	• Congruence/ • Authenticity • Address alliance ruptures with empathy • Caring, warmth, acceptance • Empathy • Collaboration
Positive expectations	• Strengthening the client's expectation of help • Ensuring successful experiences • Therapeutic rationale accepted by client/therapist		• Client's positive expectations • Provision of rationale as a change process	• Expectations and preference	
Focusing treatment					• Structured, focused treatment
Instigating change	• Provision of new information via • Self discovery • Instruction • Examples • Facilitation of emotional arousal	• Therapeutic interventions • Therapeutic realizations	• Opportunity for catharsis • Acquisition/ practice of new behaviors	• Feedback • Quality of relational interpretations • Self-disclosure	• Facilitated self-exploration • Addressing intrapersonal aspects • Address interpersonal aspects • Accurate relational interpretations • Limited relational interpretations
Responsiveness				• Managing resistance • Customizing	• Appropriate responsiveness

Note. Divisions between factors are used to demonstrate how factors were grouped to form the G-COMP factors.

observation into conceptually similar categories: (a) *alliance-building*, (b) *facilitating positive expectations*, (c) *instigating change*, (d) *focusing treatment*, and (e) *responsiveness*.

Alliance describes the emotional bond, collaboration on the tasks of treatment, and consensus on treatment goals (Bordin, 1979). Meta-analyses show that the alliance accounts for a significant portion of the variance in outcome, with mean correlations of .14–.22 for youth (McLeod, 2011; Shirk, Karver, & Brown, 2011) and .28 for adults (e.g., Horvath, Del Re, Fluckiger, & Symonds, 2011). Alliance-building behaviors that significantly predict treatment outcome that were included in the scoring manual include demonstrating positive regard (r = .27; Farber & Doolin, 2011), empathy (r = .31; Elliott, Bohart, Watson, & Greenberg, 2011), goal consensus (r = .34), and collaboration (r = .33; Tryon & Winograd, 2011).

Client expectations regarding their role and potential outcome of treatment was identified as a promising (Norcross, 2011) and probably efficacious aspect of the therapeutic relationship (Norcross, 2002). A meta-analysis has shown a weighted correlation of r = .12 for treatment outcome expectancies and outcome in adult studies (Constantino, Arnkoff, Glass, Ametrano, & Smith, 2011). Within the youth literature, parent pretreatment expectancies have been shown to predict treatment attendance and engagement (Nock & Kazdin, 2001; Nock, Ferriter, & Holmberg, 2007). There is further evidence that therapist behaviors including psychoeducation, assessing and discussing the client's beliefs about treatment, addressing skepticism, and activating beliefs that the client is being helped can change client treatment expectations (Arnkoff, Glass, & Shapiro, 2002; Swift & Callahan, 2011).

Instigating change encompasses therapist activities directed at initiating change processes; what Orlinsky and Howard (1987) referred to as "the official 'business' of seeking and giving help" (p.8) by facilitating the client's self-exploration and learning about the presenting problem and the development of skills to alleviate those problems (e.g. interpersonal skills, self-awareness). We defined instigating change as the therapist's ability to focus the client's attention upon interpersonal and/or intrapersonal psychological processes and facilitate new ways of thinking, acting, or behaving toward him/herself and others. Examples of therapist behaviors that were included in the manual to reflect instigating change are as follows: providing feedback (Claiborn, Goodyear, & Horner, 2002); self-disclosure (Hill & Knox, 2002); accurate relational interpretations (Crits-Christoph & Connolly Gibbons, 2002)[1]; and those identified by the Task Force for Empirically Based Principles of Therapeutic Change including addressing interpersonal issues, modification of behavioral, emotional, or physiological responses, facilitating self-exploration, and facilitating emotion regulation (Castonguay & Beutler, 2006).

Focusing treatment represents the therapist's ability to make efficient use of time during a session and concentrate on key themes (Follette & Greenberg, 2006; Woody & Ollendick, 2006). The Task Force on Empirically Based Principles of Therapeutic Change (Castonguay & Beutler, 2006) concluded that therapists who provide structured treatment focused on their interventions of choice are more likely to affect positive change. Research suggests that the consistent use of interventions belonging to a single approach may be more important than the specific approach itself (Luborsky, McLellan, Woody, O'Brian, & Auerbach, 1985). For the G-COMP, we included the structure and pacing of the session, continuity of themes across sessions, and key themes as indicators of focusing treatment (Woody & Ollendick, 2006).

Responsiveness to individual client needs and customization of treatment represents the final proposed global factor. The APA Division 29 Task Force (Norcross, 2002, 2011) emphasized the importance of adapting treatment to individual client needs according to resistance, preferences, culture, and religion/spirituality. We characterized responsiveness as the therapists' ability to address client motivation, particularly in the face of resistance and other disruptions in the therapeutic relationship, and flexibility in adapting interventions to meet the cultural, developmental, or immediate needs of the client.

1.1.2 | Item generation and refinement

In accord with common psychometric practice, multiple sources were sampled when drafting the items for each scale to establish their preliminary content validity (Clark & Watson, 1995). Because there were no existing stand-alone

instruments of global competence in the youth literature, we included an instrument of adherence and competence for adolescent substance abuse treatment (The Therapist Behavior Rating Scale-Competence; Hogue et al., 2008) and instruments from the adult literature (e.g., Cognitive Therapy Adherence and Competence Scale; Barber, Liese, & Abrams, 2003; Yale Adherence and Competence Scale; Carroll et al., 2000) and modified them for use in youth treatment. The initial review resulted in a large set of items ($n = 265$) that were categorized into the five G-COMP domains. Given the labor-intensive nature of rating treatment sessions, it was not possible to conduct a data-driven approach to item reduction. This initial list was reduced by eliminating items that were (a) redundant, (b) domain-specific, or (c) not easily observable. Three of the study authors (Brown, McLeod, Southam-Gerow) independently voted (yes/no) on whether to retain each item. Items with at least two affirmative responses were retained, resulting in a reduced set of items ($n = 125$).

Next, we invited therapists from a variety of settings and theoretical orientations to participate in an institutional review board-approved online survey to evaluate content validity of the items via emails sent to universities with clinical child psychology training programs as well as listservs of APA Divisions 12, 17, 29, and 53. A total of 66 participants (39.4% male, 90.9% White, 59.0% licensed, 68.0% PhD, 67.0% primarily child/adolescent therapist) rated each item according to how well the item (a) assessed therapist competence (0 = *poor item* to 6 = *excellent item*), and (b) represented all forms of youth treatment (0 = *specific to only one approach* to 6 = *common to all approaches*).

Items with mean quality and commonality ratings greater than four (greater than the midpoint) were retained for further consideration, resulting in a pool of 44 items. The 44 items were given to therapists who had expertise in a variety of therapeutic models (i.e., five Ph.D. level psychologists including two developers of youth CBT treatment manuals, and three senior faculty with experience providing and supervising psychoanalytic, psychodynamic, and interpersonal therapies) for feedback. We reviewed the feedback and made final decisions to retain, combine, or remove items, which reduced the item set from 44 to 24.

Two coders piloted the 24-item G-COMP instrument with 16 recordings, which led to the elimination of five items due to poor inter-rater reliability assessed with ICC(2,2) < .40. The final version of the G-COMP had 19 items organized into two levels. Level-1 items included five superordinate "macro-level" items representing the five main domains: (a) alliance-building, (b) facilitating positive expectations, (c) instigating change, (d) focusing treatment, and (e) responsiveness. Level-2 items included 14 subordinate "micro-level" items that define specific classes of therapist behaviors that contribute to Level-1 domains.

1.1.3 | Scoring strategy

We adopted a scoring strategy used in exemplar competence coding systems used in youth (Hogue et al., 2008) and adult (Carroll et al., 2000) treatment. In making competence ratings, coders are asked to consider the quality of the therapists' attempts to manage the therapeutic relationship and/or promote change (skillfulness) and their timing and appropriateness for the given client and situation (responsiveness) on a 7-point Likert-type competence scale ranging from 1 (*very poor*) to 7 (*excellent*).

1.1.4 | Scoring manual development

The scoring manual (Brown, McLeod, Southam-Gerow, Bair, & Wheat, 2012) was developed following the exemplar coding manuals from the youth (Hogue et al., 2008) and adult (Carroll et al., 2000) treatment field. For each item, coders were asked to consider the therapist's expertise in promoting change, clarity of communication, appropriate timing of actions, and ability to assess and adequately respond to the client's developmental level and current functioning. The manual includes guidance on making ratings when parents or caregivers are present. The manual includes overall and item-specific guidelines for each anchor of the scale, detailed descriptions of each item, examples of behaviors consistent with each item, examples of behaviors that might represent each scaling anchor, and guidance on making distinctions between related items.

1.2 | Psychometric evaluation

1.2.1 | Data sources and participants

Participants in the study included 68 youth who met diagnostic criteria for a principal anxiety disorder according to *Diagnostic and Statistical Manual of Mental Disorders* criteria (American Psychiatric Association, 1994) from two randomized controlled trials (RCTs), conducted by separate research groups. Recorded treatment sessions were coded and the resulting data were analyzed for this study. Inclusion criteria for the recordings were (a) a minimum of two audible sessions and (b) received treatment from a single therapist (for more details, see Kendall, Hudson, Gosch, Flannery-Schroeder, & Suveg, 2008, and Southam-Gerow et al., 2010). See Table 2 for descriptive statistics of youth. multiple institutional review boards approved the study procedures.

A total of 16 therapists (12.5% male; 81.3% White; 6.3% Latino, 6.3% Asian/Pacific Islander, and 6.3% did not report) participated in the first RCT (ICBT; Kendall et al., 2008). These therapists were either clinical psychology doctoral trainees or licensed clinical psychologists. The second RCT, the Youth Anxiety Study (YAS; Southam-Gerow et al., 2010), included therapists who were clinic employees, who volunteered to participate in the study, and randomly assigned to treatment groups. Therapists assigned to YAS-ICBT ($n = 13$; 15.4% male) were 53.8% White, 15.4% Latino, 15.4% Asian/Pacific Islander, and 15.4% mixed/other. Professional composition of YAS-ICBT therapists was 30.8% social workers, 23.1% master's-level psychologists, 15.3% doctoral-level psychologists, and 30.8% reported "other" degree.

1.2.2 | ICBT

Therapists in ICBT and YAS-ICBT delivered Coping Cat, an ICBT program designed for youth diagnosed with anxiety disorders (Kendall & Hedtke, 2006). Coping Cat emphasizes anxiety management skills training (e.g., cognitive restructuring, relaxation), exposure, and regular homework across 16 sessions (14 conducted individually with the youth and 2 conducted with parents only). In both studies, the therapists were trained using the same procedure: treatment manual, training workshop, and ongoing supervision with a model expert.

1.3 | Instruments used for validity analyses

The CBT for Anxiety in Youth Competence Scale (CBAY-C; McLeod et al., 2016) is a 25-item observational instrument that assesses three areas of technical competence: (a) *standard* comprises five items that represent interventions found in most CBT sessions (e.g., Agenda Setting, Homework Review); (b) *model* comprises 12 items that assess model-specific content (i.e., Psychoeducation-Anxiety, Emotion Education, Fear Ladder, Relaxation, Cognitive-Anxiety, Problem Solving, Self-Reward, Coping Plan, Exposure Preparation, Exposure, Exposure Debrief, Maintenance); and (c) *delivery* comprises six items that measure how model items are delivered (e.g., Didactic Teaching, Rehearsal) and two Global ratings.

In making competence ratings, coders are asked to watch an entire session and make ratings on a 7-point Likert-type competence scale with the following anchors: 1 = *very poor*, 3 = *acceptable*, 5 = *good*, and 7 = *excellent*. The CBAY-C total scale, subscale, and item scores have demonstrated evidence of construct validity (McLeod et al., 2016). However, the subscale scores were highly correlated in this sample ($r = .86$, $p < .001$), so we used the total scale. The total scale score was produced by taking the highest scoring item for each session (see McLeod et al., 2016; Smith et al., 2016). For the present sample, the average item inter-rater reliability, ICC(2,2) = 0.69 ($SD = 0.09$).

CBT Adherence Scale for Youth Anxiety (CBAY-A; Southam-Gerow et al., 2016) is a 22-item observer-rated instrument of model-specific adherence to ICBT for youth anxiety that parallels the content of the CBAY-C. The CBAY-A assesses the same three areas as the CBAY-C (Standard, Model, and Delivery) rated on a 7-point extensiveness scale (Hogue, Liddle, & Rowe, 1996), ranging from 1 (*not at all)* to 7 (*extensively)*. The present study used the total scale score, which was generated by using the item with the highest score for each recording. The CBAY-A item and subscale scores have demonstrated evidence of construct validity and discriminated between therapists delivering ICBT across research and community settings from therapists delivering usual care (Southam-Gerow et al., 2016). For the present sample, the average item inter-rater reliability, ICC(2,2,) = 0.75 (standard deviation [SD] = 0.16).

TABLE 2 Youth descriptive data and comparisons across Groups

Variable	ICBT (N = 51)	*M (SD)* or % YAS-ICBT (N = 17)	*F* or Chi Square [a]
Age	10.36 (1.90)	11.32 (2.32)	1.72
Sex			
Male	60.8%	29.4%	5.04*
Race/ethnicity			18.97**
White	86.3%	41.2%	
African American	9.8%	–	
Latino	2.0%	17.6%	
Mixed/other	2.0%	5.9%	
Not reported	–	35.3%	
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)	2.87**
Primary diagnoses			
GAD	37.3%	5.9%	
SAD	29.4%	35.3%	
SOP	33.3%	23.5%	
SP	–	35.3%	
Family income			7.92**
Up to 60k per year	35.3%	70.6%	
Number of sessions	15.92 (1.43)	16.82 (5.02)	0.73
Weeks in treatment	19.52 (3.97)	26.38 (10.41)	2.65**
	Adj. M		
G-COMP	**ICBT n = 532**	**YAS-ICBT n = 212**	***t*** [b]
Alliance building	4.68	4.38	6.03**
Positive expectancies	3.97	3.42	9.69**
Focusing treatment	4.91	4.06	13.27**
Instigating change	4.70	3.76	12.87**
Responsiveness	4.50	4.10	7.19**

Note. ICBT = individual cognitive-behavioral therapy delivered in Kendall et al. (2008) study; YAS-ICBT = ICBT delivered in Youth Anxiety Study; CBCL = Child Behavior Checklist; GAD = generalized anxiety disorder; SAD = separation anxiety disorder; SOP = social phobia; SP = specific phobia.

[a]Analysis of variance was conducted with continuous variables whereas chi-square analyses were conducted with categorical variables.

[b]Least squares means difference test was conducted.

Therapy Process Observational Coding System for Child Psychotherapy-Alliance scale (TPOCS-A; McLeod & Weisz, 2005) is an observational instrument designed to assess the quality of the client-therapist alliance in youth treatment. The TPOCS-A comprises 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 inter-rater reliability ranging from .48 to .80 (mean [*M*] ICC = .67) and internal consistency ranging from .91 to .95 ($M\ \alpha = .92$). Scores on the TPOCS-A 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 treatment outcomes (Chiu, McLeod, Har, & Wood, 2009; Liber et al., 2010; McLeod & Weisz, 2005). Inter-rater reliability, ICC(2,2), for the TPOCS-A in the present sample was .82 and internal consistency was .81.

Therapeutic Alliance Scale for Children (TASC; Shirk & Saiz, 1992) is a youth-report instrument of alliance. Items are scored on a 4-point scale ranging from 1 (*not true at all*) to 4 (*very true*). The TASC has evidenced convergent validity with other alliance instruments (Accurso, Hawley, & Garland, 2013; Fjermestad et al., 2012) and predictive validity in relation to outcome (e.g., Hawley & Weisz, 2005). In YAS, the seven-item version (Shirk & Saiz, 1992) was administered, whereas the revised 12-item version was administered in Kendall et al. (2008). Six items had identical wording, so scores were based on a sum of those six items. These six TASC items focus on the bond dimension of the alliance (e.g., "I felt like my therapist was on my side and tried to help me"). For analyses, we used TASC scores collected at posttreatment. [2] The internal consistency for the six items used was $\alpha = .81–.86$ across samples.

Child Involvement Rating Scale (CIRS; Chu & Kendall, 2004) is a six-item instrument that assesses aspects of positive and negative youth involvement in treatment. Coders view entire sessions and then rate items on a 6-point scale ranging from 0 (*not at all*) to 5 (*a great deal*). Previously, the CIRS has demonstrated inter-rater reliability (ICC) from .61 to .90 and an internal consistency of .73 (Chu & Kendall, 2004, 2009; Hudson et al., 2014). Inter-rater reliability, ICC(2,2), for the CIRS in this study was .78; internal consistency was .85.

1.4 | Study observational coding procedures

1.4.1 | Coders

The coding teams comprised eight doctoral students in clinical psychology: Two students (*M age* = 27.00 years, *SD* = 0.00; 100.0% female; 100.0% White) coded the G-COMP; two students (*M age* = 27.00 years, *SD* = 4.20; 50.0% female; 50.0% White, 50.0% Latina) coded the CBAY-A and CIRS; two students (*M age* = 24.00 years, *SD* = 1.40; 100.0% female; 50.0% White, 50.0% Asian-American) coded the TPOCS-A; and four students coded the CBAY-C (*M age* = 28.00 years, *SD* = 2.70; 75.0% female; 50.0% White, 50.0% Latina).

1.4.2 | 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 (second and third author), and engaged in exercises designed to expand understanding of each item. Second, coders engaged in coding and results were discussed in weekly meetings. Last, coders independently coded 32 recordings and reliability was assessed against master codes produced by the second and third authors. To be certified for independent coding, each coder had to demonstrate reliability of ICC > .59 on each item (Cicchetti, 1994).

1.4.3 | Assignment and coding of sessions

All treatment 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) they were shorter than 15 minutes, (b) less than 15 minutes 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) and 74.1% YAS-ICBT (n = 212). There was no significant difference between groups in terms of the percent of sessions coded, $t(66) = 1.85$, $p = .069$, nor was there a difference in the percent of sessions coded from the first and second half of treatment (first half = 67.6%; second half 67.9%), $t(67) = 0.07$, $p = .95$.

1.5 | Data analysis

1.5.1 | Inter-rater reliability

Intraclass correlation coefficients, ICC(2,2), based on a two-way random effects model, were calculated across coders for each G-COMP item to examine inter-rater reliability (Shrout & Fleiss, 1979). We considered ICC

values below .40 to reflect "poor" agreement, ICCs from .40 to .59 to reflect "fair" agreement, ICCs from .60 to .74 to reflect "good" agreement, and ICCs .75 and higher to reflect "excellent" agreement (Cicchetti, 1994).

1.5.2 | Item reduction

To determine the most appropriate, and parsimonious, level of analysis (e.g., Level-1 items vs. average of Level-2 items), we examined correlations between the Level-1 items (e.g., Alliance Building) and the sum of the corresponding Level-2 items.

1.5.3 | Variance components analysis

Variance components analysis of the five G-COMP Level-1 items (Alliance, Positive Expectancies, Focusing Treatment, Instigating Change, Responsiveness) was conducted using a random effects linear mixed-effects model (MIXED), with restricted maximum likelihood estimation (REML) in SPSS to estimate the sources of variance around G-COMP scores. Variance components were calculated for the following factors: (a) Setting (i.e., research for ICBT or community for YAS-ICBT); (b) Therapist (nested within setting); (c) Client (nested within setting, therapist); (d) Time (i.e., weeks since intake nested within client, therapist, setting); and (e) Coder. Estimates of variance were transformed into proportions of the total variance. Mean level differences between settings on the G-COMP item scores were tested using adjusted least square means scores with SAS/STAT Software 9.4 using with a Bonferroni adjusted alpha of $p < .01$.

1.5.4 | Construct validity of the G-COMP scores

Next, we evaluated the extent to which the G-COMP item scores demonstrated evidence of construct validity by investigating the magnitude of the association between the G-COMP item scores and scores on other treatment process instruments, including observational (TPOCS-A) and youth-report (TASC) alliance instruments, an observational client involvement instrument (CIRS), an observational CBT-competence instrument (CBAY-C Total scale), and an observational instrument of ICBT adherence (CBAY-A Total scale). We hypothesized that inter-item correlations on the G-COMP would be medium to large because these items were designed to assess moderately related components of global therapist competence. We hypothesized that correlations would be positive across the instruments; however, given the absence of gold standard instruments of global therapist competence to which to compare the G-COMP, we made no specific hypotheses about the magnitude.

2 | RESULTS

2.1 | Descriptive statistics and inter-rater reliability

Means of the G-COMP items generally fell around the midpoint of the 7-point scale and the full range of the scores was represented for each item. ICCs for the G-COMP Level-2 item scores ranged from .49 to .81 ($M = 0.65$, $SD = 0.10$), with five items in the "fair" range, seven in the "good" range, and two in the "excellent" range. ICCs were higher among the Level-1 items ($M = 0.70$, $SD = 0.07$), with all ICCs > .60. See Table 3 for details.

2.2 | Item reduction

The correlations between each Level-1 item score and the sum of the corresponding Level-2 item scores ranged from 0.92 to 0.97, indicating redundancy (r > .70; Kline, 1979). These results suggest that the Level-1 scores represent a more parsimonious scoring approach for the G-COMP. Thus, we used only the five Level-1 item scores in subsequent analyses.

TABLE 3 Interrater reliability, means, and standard deviations for G-COMP level-1 and level-2 items

	Total N = 744			
Item Description	**ICC**	***M***	**(*SD*)**	**Range**
I. Alliance building	0.67	4.59	(0.75)	1–7
1. Understanding	0.53	4.10	(0.69)	1–7
2. Positive regard	0.71	4.75	(0.87)	1–7
3. Client's perspective	0.49	4.35	(0.69)	2–7
4. Collaboration	0.59	4.62	(0.89)	1–7
II. Positive expectancies	0.70	3.82	(0.89)	1–7
5. Treatment expectancies	0.68	3.60	(1.14)	1–7
6. Therapist credibility	0.61	4.20	(0.63)	1.5–7
7. Client self-efficacy	0.72	3.46	(1.13)	1–7
III. Focusing treatment	0.74	4.66	(1.05)	1–7
8. Structure/Pace	0.73	4.65	(1.14)	1–7
9. Continuity	0.53	3.88	(0.81)	1.5–6.5
10. Key Themes	0.74	4.98	(1.13)	1–7
IV. Instigating change	0.79	4.43	(1.22)	1–7
11. Change strategies	0.76	4.50	(1.16)	1–7
12. Active participation	0.81	4.37	(1.32)	1–7
V. Responsiveness	0.61	4.39	(0.83)	1–7
13. Motivation	0.57	4.37	(0.81)	1–7
14. Flexibility	0.65	4.31	(0.83)	1–7

Note. M = mean; SD = standard deviation; ICC = intraclass correlation coefficient; G-COMP = Global Therapist Competence Scale for Youth Psychosocial Treatment.
ICC calculated as (2,2) two-way random effects model of two coders.

TABLE 4 Variance components for G-COMP ratings

	Variance Components					
G-COMP Scale	**Setting**	**Therapist**	**Client**	**Time**	**Coder**	**Residual**
Alliance building	3%	19%	4%	22%	1%	52%
Positive expectancies	9%	11%	4%	29%	<1%	47%
Focusing treatment	13%	21%	3%	26%	1%	36%
Instigating change	17%	15%	3%	35%	<1%	31%
Responsiveness	14%	26%	3%	25%	<1%	32%

Note. G-COMP = Global Therapist Competence Scale for Youth Psychosocial Treatment. Variance component estimates represent the portion of variance that is attributed to each source of variance.

2.3 | Variance components analysis

Table 4 contains the results of the variance components analysis for each Level-1 item score. Most of the nonresidual variance was accounted for by time (22.0% –35.0%), therapist (11.0%–26.0%), and setting (3.0%–17.0%). Neither coder (≤ 1.0%) nor client (≤ 4.0%) accounted for much variance. Next, we tested for differences between settings in G-COMP Level-1 item scores using adjusted least square means (Table 2). Post hoc tests, using Bonferroni corrections to account for multiple comparisons, revealed that means for all Level-1 items were significantly higher for the ICBT group than the YAS-ICBT (*p*s < .001).

TABLE 5 Correlations among G-COMP level-1 items and process instruments

	n	1	2	3	4	5	6	7[a]	8	9
1. G-COMP alliance building	744									
2. G-COMP positive expectancies	744	.43**								
3. G-COMP focusing treatment	744	.45**	.53**							
4. G-COMP instigating change	744	.52**	.37**	.75**						
5. G-COMP responsiveness	744	.60**	.38**	.48**	.55**					
6. TPOCS-A	660	.44**	.26**	.27**	.43**	.18**				
7. TASC[a]	64	.27*	.21	.23	.17	.22	.28*			
8. CIRS	661	.32**	.10**	.13**	.27**	.08*	.77**	.21		
9. CBAY-C total	663	.26**	.24**	.53**	.52**	.26**	.25**	.28*	.21**	
10. CBAY-A total	744	.27**	.25**	.55**	.63**	.26**	.38**	.21	.23**	.65**

Note. Coefficients represent correlation with process scores from the last observation.
G-COMP = Global Therapist Competence Scale for Youth Psychosocial Treatment; CBAY-C = CBT for Youth Anxiety Competence Scale; TPOCS-A = Therapy Process Observational Coding System for Child Psychotherapy-Alliance scale; TASC = Therapeutic Alliance Scale for Children; CIRS = Child Involvement Rating Scale; CBAY-A = CBT Adherence Scale for Youth Anxiety.
[a]TASC assessed at end of treatment.
*$p < .05$. **$p < .01$.

2.4 | Construct validity of G-COMP item scores

2.4.1 | Inter-item correlations

As seen in Table 5, inter-item correlations among the five G-COMP items ranged from .37 to .75 (*p*s < .001), suggesting the item scores assess related yet distinct aspects of global competence with one exception: the correlation between scores on the Focusing Treatment and Instigating Change items ($r = .75$, $p < .001$).

2.4.2 | Correlations between G-COMP and measures of alliance and involvement

The correlation between the G-COMP item scores and scores on the TPOCS-A (observer-rated alliance) ranged from $r = .18$ –.44. The TPOCS-A was most strongly correlated with the G-COMP's Alliance ($r = .44$, $p < .001$) and Instigating Change ($r = .43$, $p < .001$) items. The correlations between the last observation of G-COMP and the TASC (end-of-treatment youth-reported measure of alliance), ranged from *r*s = .17–.27, with the highest, and only significant correlation with the G-COMP Alliance item. The correlations between the G-COMP and the CIRS (observer-rated child involvement) were small to moderate, ranging from .08 to .32, (*p*s < .05), with the highest correlations between G-COMP Alliance and Instigating Change.

2.4.3 | Correlations between G-COMP and measures of CBT-competence and adherence

The strongest observed correlations with the G-COMP items were with instruments of technical delivery: CBT-adherence (CBAY-A; *r*s = .25 - .63) and CBT-competence (CBAY-C; *r*s = .24 - .53). Specifically, the correlations with the G-COMP Instigating Change item (*r*s = .63 and .52, *p*s < .001, for CBAY-A and CBAY-C, respectively) and with the Focusing Treatment item (*r*s = .55 and .53, *p*s < .001, respectively) were the highest.

3 | DISCUSSION

Previous research has found inconsistent and weak associations between therapist competence and treatment outcomes, potentially because of the lack of conceptual distinction between global and technical therapist competence and adherence (Barber et al., 2006; Barber et al., 2007; Hogue et al., 2008). The G-COMP is, to our knowledge, the first stand-alone measure of global therapist competence for youth treatment. The G-COMP assesses therapist

competence across five domains: (a) alliance building, (b) promoting positive expectancies, (c) focusing treatment, (d) instigating change, and (e) responsiveness. Our results supported the notion that the five domains were best represented by the single Level-1 item scores, suggesting that future research could simply use the five Level-1 items. However, this may indicate that coders may have difficult making fine grain distinctions between some of the categories that comprise the broader global domains (cf. Hogue et al., 2008; McLeod et al., 2016).

We found that independent coders can achieve acceptable inter-rater reliability on the G-COMP items. Inter-rater reliability of the Level-1 items was in the good to excellent range (ICC ≥ .60; Cicchetti, 1994), higher than previous estimates of global competence (ICC .48 to .56; e.g., Hogue et al., 2008), and consistent with pervious findings (ICC .71 to .85; e.g., Carroll et al., 2000). This result was corroborated by the negligent variance component associated with coder (< 1%). This brings some preliminary evidence to bear on the question of whether global competence can be operationalized and coded reliably (Barber et al., 2007).

The variance components analysis resulted in several findings. Client differences accounted for very little variance (between 3.0% and 4.0%), suggesting that scores across the five global domains did not vary significantly at the client level. This is desirable given that the G-COMP was designed to focus on therapist competence and is consistent with some past work (Barber, Foltz, Crits-Christoph, & Chittams, 2004), although counter to other findings (Hogue et al., 2008). Time in treatment accounted for 22.0%–35.0% of the variance in G-COMP scores, consistent with previous findings (e.g., Barber et al., 2004; McLeod et al., 2016). This suggests that there may be systematic differences in scores on the competence domains from session to session. The variance associated with therapist (11%–26%) suggests that scores on the five domains differ across therapists. This is a positive sign for an instrument of therapist competence and is consistent with some past work (Barber et al., 2004; Hogue et al., 2008). The relatively low number of clients per therapist, particularly in the YAS-ICBT, may have reduced the variance to be accounted for by therapists in our sample.

Setting accounted for a modest amount of variance (3.0%–17.0%) and therapists in the research setting were rated higher than therapists in the community setting. The difference is consistent with studies reporting significant differences in treatment adherence and technical competence between effectiveness and efficacy trials (e.g., McLeod, Smith, Southam-Gerow, Weisz, & Kendall, 2015; Weisz et al., 2009). Although the research and community settings did not differ in *number* of sessions, clients in the community setting spent more *weeks* in treatment, which may have affected treatment continuity. Therapists in the two settings differed in their training experience and background during the study, which may have affected therapist competence (Henggeler, Schoenwald, Liao, Letourneau, & Edwards, 2002). Therapists in the research setting were more experienced in the use of the specific treatment manual than the therapists in the community setting, who were newly trained. There is evidence that therapists newly trained in a manual-guided treatment may exhibit a temporary decline in interpersonal skillfulness as they struggle to internalize the new interventions (Henry, Strupp, Butler, Schacht, & Binder, 1993) and it is possible that the G-COMP is sensitive to these differences.

Inter-item correlations of the Level-1 items were all significant. This suggests that therapists who exhibit competency in one domain may be more likely to exhibit competency in another, which is consistent with prior work (e.g., Carroll et al., 2000; Hogue et al., 2008). The magnitude of the correlation coefficients between Focusing Treatment and Instigating Change ($r = .75$) suggests significant conceptual overlap. The items may be redundant ($r > .70$; Kline, 1979), or the items may co-vary within sessions (e.g., therapists who demonstrate competence in instigating change may also be skillful at keeping treatment focused on specific goals). Future studies are needed to further examine the discriminative validity of scores on these two items.

The pattern of correlations between on the G-COMP, alliance, and client involvement instruments support the discriminant validity of the item scores. Conceptual models emphasize that therapist competence promotes client involvement in part by strengthening the alliance (Fjermestad et al., 2016). Correlations between scores on the G-COMP item, the alliance instruments (both observer and self-report), and the client involvement instrument were positive. The finding that the G-COMP alliance item was differentially correlated with the client-report alliance instrument was particularly promising.

We also examined relations between the G-COMP items and the instruments designed to assess both model-specific therapist competence and therapeutic interventions. The range of correlations between scores on the G-COMP and instruments of technical competence and adherence (rs = .24–.63) were less than what has been observed in other studies that have assessed common competence with items embedded in technical competence/adherence scales (e.g., *r*s = .79 –.91, Stallard et al., 2014). The use of independent coders likely contributed to improved conceptual distinction. The pattern of high, but not redundant, positive correlations with a technical competence scale suggests that therapists skilled at implementing manual-guided ICBT for youth anxiety disorders were also likely to exhibit similar levels of global competencies, consistent with prior research (e.g., Carroll et al., 2000; Hogue et al., 2008).

3.1 | Limitations

The results must be considered in light of some limitations. Though observer-rated instruments have strengths, they also limit the scope of skills that can be assessed to those that can be observed. To provide a more thorough assessment of therapist competence, a multimethod approach would be needed to tap various aspects of this broader definition of competence. Another potential limitation is our use of advanced graduate student coders. Although we found good inter-rater reliability, and previous studies have found good inter-rater reliability between expert and nonexpert judges of general therapeutic skills (Barber & Crits-Christoph, 1996), future research is needed to examine the predictive validity of scores with treatment outcome.

Another limitation is that we examined the G-COMP only in the context of ICBT for anxiety disorders, and thus we are not able to determine the extent to which G-COMP items are truly global across other theoretical orientations and/or client problems. The G-COMP was based heavily on the empirical work of the APA's Division 29 Task Force on Empirically Supported Relationships (Norcross, 2002, 2011), with its emphasis on theorietical neutrality. We would expect similar results from short-term manual-guided treatments, such as psychodynamic treatment for anxiety disorders (Leichsenring et al., 2013), but this should be examined in future research. Finally, we examined the G-COMP in the context of ICBT for anxiety disorders; thus, results may not generalize to the treatment of other disorders. It is possible that different profiles of skills may emerge in the treatment of externalizing problems. Future studies are needed to examine the properties of the G-COMP in treatments for other diagnoses.

4 | CONCLUSION

This study marks an important first step in developing an instrument of global competence for youth treatment. The present study indicates that graduate students can code the G-COMP items reliably. Moreover, evidence provides some support for the construct validity of the item scores (Foster & Cone, 1995), suggesting that they are distinct from other treatment process domains. This suggests that the G-COMP may be used in conjunction with other treatment process instruments to identify the breadth of treatment and relationship elements that are associated with meaningful clinical outcomes. This line of research will be critical in developing evidence-based training and practice guidelines for youth treatment.

NOTES

[1] In the second edition of the Task Force on Evidence-Based Psychotherapy Relationships, formerly the Task Force on Empirically Supported Relationships, (Norcross, 2011), published during the G-COMP coding, self-disclosure and transference interpretations were no longer included because of sufficient evidence.

[2] TASC was collected only at posttreatment in YAS (Southam-Gerow et al., 2010).

ORCID

Ruth C. Brown http://orcid.org/0000-0001-8614-1645

REFERENCES

Accurso, E. C., Hawley, K. M., & Garland, A. F. (2013). Psychometric properties of the Therapeutic Alliance Scale for Caregivers and Parents. *Psychological Assessment*, *25*, 244–252.

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders (4th ed.)*. Washington, DC: Author.

American Psychological Association. (2015). Standards of accreditation for health service psychology. Retrieved from www.apa.org/ed/accreditation/about/policies/standards-of-accreditation.pdf

Arnkoff, D., Glass, C., & Shapiro, S. (2002). Expectations and preferences. In J. C. Norcross (Ed.), *Psychotherapy Relationships That Work: Therapist Contributions and Responsiveness to Patients* (pp. 335–356). London: Oxford University Press.

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*, 29–41.

Barber, J. P., Gallop, R., Crits-Christoph, P., Frank, A., Thase, M. E., Weiss, R. D., … Gibbons, M. B. (2006). The role of therapist adherence, therapist competence, and alliance in predicting outcome of individual drug counseling: Results from the National Institute Drug Abuse Collaborative Cocaine Treatment Study. *Psychotherapy Research*, *16*(2), 229–240.

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

Barber, J. P., Sharpless, B. A., Klostermann, S., & McCarthy, K. S. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology: Research and Practice*, *38*, 493–500.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice*, *16*(3), 252–260.

Boswell, J. F., Gallagher, M. W., Sauer-Zavala, S. E., Bullis, J., Jack, M., Shear, M. K., … York, N. (2013). Patient characteristics and variability in adherence and competence in cognitive-behavioral therapy for panic disorder. *Journal of Consulting & Clinical Psychology*, *81*, 443–454.

Brown, R. C., McLeod, B. D., Southam-Gerow, M. A., Bair, C. E., & Wheat, E. (2012). Scoring manual for the Common Factor Therapist Competence Scale. Unpublished instrument.

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*, 225–238.

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

Castonguay, L. G., & Beutler, L. E. (2006). Principles of therapeutic change: A task force on participants, relationships, and techniques factors. *Journal of Clinical Psychology*, *62*, 631–638.

Chiu, A. W., McLeod, B. D., Har, K., & Wood, J. J. (2009). Child–therapist alliance and clinical outcomes in cognitive behavioral therapy for child anxiety disorders. *Journal of Child Psychology and Psychiatry*, *50*, 751–758.

Chu, B. C., & Kendall, P. C. (2004). Positive association of child involvement and treatment outcome within a manual-based cognitive-behavioral treatment for children with anxiety. *Journal of Consulting and Clinical Psychology*, *72*, 821–829.

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*, 736–754.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284–290.

Claiborn, C. D., Goodyear, R. K., & Horner, P. A. (2002). Feedback. In J. C. Norcross (Ed.), *Psychotherapy Relationships That Work* (pp. 217–233). New York: Oxford University Press.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309–319.

Constantino, M. J., Arnkoff, D. B., Glass, C. R., Ametrano, R. M., & Smith, J. Z. (2011). Expectations. *Journal of Clinical Psychology*, *67*(2), 184–192. https://doi.org/10.1002/jclp.20754

Crits-Christoph, P., & Gibbons, M. B. C. (2002). Relational interpretations. In J. C. Norcross (Ed.), *Psychotherapy relationships that work* (pp. 285–300). New York: Oxford University Press.

Elliott, R., Bohart, A. C., Watson, J. C., & Greenberg, L. S. (2011). Empathy. *Psychotherapy*, *48*(1), 43–49.

Farber, B. A., & Doolin, E. M. (2011). Positive regard. *Psychotherapy*, *48*(1), 58–64.

Fjermestad, K. W., McLeod, B. D., Heiervang, E. R., Havik, O. E., Ost, L., & 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 and Adolescent Psychology*, *41*, 246–254.

Fjermestad, K. W., McLeod, B. D., Tully, C. B., & Liber, J. M. (2016). Therapist characteristics and interventions: Enhancing alliance and involvement with youth. In S. Maltzman (Ed.), *Oxford handbook of treatment processes and outcomes in counseling psychology* (pp. 97–116). New York: Oxford University Press.

Follette, W. C., & Greenberg, L. S. (2006). Techniques factors in treating dysphoric disorders. In L. G. Castonguay & L. E. Beutler (Eds.), *Principles of therapeutic change that work* (pp. 83–109). New York: Oxford University Press.

Foster, S. L., & Cone, J. D. (1995). Validity issues in clinical assessment. *Psychological Assessment*, *7*, 248–260.

Frank, J. D. (1971). Therapeutic factors in psychotherapy. *American Journal of Psychotherapy*, *25*, 350–361.

Grencavage, L. M., & Norcorss, J. C. (1990). Where are the commonalities among the therapeutic common factors? *Professional Psychology: Research and Practice*, *21*, 372–378.

Hawley, K. M., & Weisz, J. R. (2005). Youth versus parent working alliance in usual care: Distinctive associations with retention, satisfaction, and treatment outcome. *Journal of Clinical Child and Adolescent Psychology*, *34*, 117–128.

Henggeler, S. W., Schoenwald, S. K., Liao, J. G., Letourneau, E. J., & Edwards, D. L. (2002). Transporting efficacious treatments to field settings: The link between supervisory practices and therapist fidelity in MST programs. *Journal of Clinical Child and Adolescent Psychology*, *31*, 155–167.

Henry, W. P., Strupp, H. H., Butler, S. F., Schacht, T. E., & Binder, J. L. (1993). Effects of training in time-limited dynamic psychotherapy: Changes in therapist behavior. *Journal of Consulting and Clinical Psychology*, *61*(3), 434–440.

Hill, C. E., & Knox, S. (2002). Self-disclosure. In J. C. Norcross (Ed.), *Psychotherapy relationships that work* (pp. 255–265). New York: Oxford.

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

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

Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy*, *48*, 9–16.

Hudson, J. L., Kendall, P. C., Chu, B. C., Gosch, E., Martin, E., Taylor, A., & Knight, A. (2014). Child involvement, alliance, and therapist flexibility: Process variables in cognitive-behavioural therapy for anxiety disorders in childhood. *Behaviour Research and Therapy*, *52*, 1–8.

Institute of Medicine. (2015). Psychological interventions for mental and substance use disorders: A framework for establishing evidence-based standards. Washington, DC: National Academies Press.

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

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*, 282–297.

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

Leichsenring, F., Salzer, S., Beutel, M. E., Herpertz, S., Hiller, W., Hoyer, J., … Ritter, V. (2013). Psychodynamic therapy and cognitive-behavioral therapy in social anxiety disorder: A multicenter randomized controlled trial. *American Journal of Psychiatry*, *170*, 759–767.

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*, 172–186.

Luborsky, L., McLellan, A. T., Woody, G. E., O'Brien, C. P., & Auerbach, A. (1985). Therapist success and its determinants. *Archives of General Psychiatry*, *42*, 602–611.

McLeod, B. D. (2011). Relation of the alliance with outcomes in youth psychotherapy: A meta-analysis. *Clinical Psychology Review*, *31*, 603–616.

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*, 314–325.

McLeod, B. D., Southam-Gerow, M. A., Rodriguez, A., Quinoy, A., Arnold, C., Kendall, P. C., & Weisz, J. R. (2016). Development and initial psychometrics for a therapist competence instrument for CBT for youth anxiety. *Journal of Clinical Child and Adolescent Psychology*, Advance online publication

McLeod, 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*, 323–333.

Nock, M. K., Ferriter, C., & Holmberg, E. (2007). Parent beliefs about treatment credibility and effectiveness: Assessment and relation to subsequent treatment participation. *Journal of Child and Family Studies*, *16*, 27–38.

Nock, M. K., & Kazdin, A. E. (2001). Parent expectancies for child therapy: Assessment and relation to participation in treatment. *Journal of Child and Family Studies*, *10*, 155–180.

Norcross, J. C. (2002). *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients*. New York: Oxford University Press.

Norcross, J. C. (Ed.). (2011). *Psychotherapy relationships that work: Evidence-based responsiveness*(2nd ed.). New York: Oxford University Press.

Orlinsky, D. E., & Howard, K. I. (1987). A generic model of psychotherapy. *Journal of Integrative & Eclectic Psychotherapy*, *6*, 6–27.

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*, 347–354.

Shirk, S. R., & Saiz, C. C. (1992). Clinical, empirical, and developmental perspectives on the therapeutic relationship in child psychotherapy. *Development and Psychopathology*, *4*, 713–728.

Shirk, S. R., Karver, M. S., & Brown, R. (2011). The alliance in child and adolescent psychotherapy. *Psychotherapy*, *48*, 17–24.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428.

Southam-Gerow, M. A., McLeod, B. D., Arnold, C. A., Rodriguez, A., Cox, J. R., … Kendall, P. C. (2016). Initial development of a treatment adherence measure for cognitive-behavioral therapy for child anxiety. *Psychological Assessment*, *28*, 70–80.

Southam-Gerow, M. A., Weisz, J. R., Chu, B. C., McLeod, B. D., Gordis, E. B., & Connor-Smith, J. K. (2010). Does CBT for youth anxiety outperform usual care in community clinics?: An initial effectiveness test. *Journal of the American Academy of Child and Adolescent Psychiatry*, *49*, 1043–1052.

Spruill, J., Rozensky, R. H., Stigall, T. T., Vasquez, M., Bingham, R. P., & De Vaney Olvey, C. (2004). Becoming a competent clinician: Basic competencies in intervention. *Journal of Clinical Psychology*, *60*, 741–754.

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*(03), 269–282.

Swift, J. K., & Callahan, J. L. (2011). Decreasing treatment dropout by addressing expectations for treatment length. *Psychotherapy Research*, *21*, 193–200.

Tryon, G. S., & Winograd, G. (2011). Goal consensus and collaboration. *Psychotherapy*, *48*(1), 50–57.

Vallis, T. M., Shaw, B. F., & Dobson, K. S. (1986). The Cognitive Therapy Scale: Psychometric properties. *Journal of Consulting and Clinical Psychology*, *54*, 381–385.

Woody, S. R., & Ollendick, T. H. (2006). Technique factors in treating anxiety disorders. In L. G. Castonguay & L. E. Beutler (Eds.), *Principles of therapeutic change that work* (pp. 167–186). New York: Oxford University Press.

How to cite this article: Brown RC, Southam-Gerow MA, McLeod BD, et al. The global therapist competence scale for youth psychosocial treatment: Development and initial validation. *J Clin Psychol*. 2018;74:649–664. https://doi.org/10.1002/jclp.22537