ORIGINAL ARTICLE

CrossMark

The Factor Structure of the Cognitive Therapy Rating Scale (CTRS) in a Sample of Community Mental Health Clinicians

Nicholas W. Affrunti[1] · Torrey A. Creed[1,2]

Abstract

Treatment fidelity is an essential outcome of implementation research. The gold standard measure for Cognitive-Behavioral Therapy (CBT) fidelity is the Cognitive Therapy Rating Scale (CTRS). Despite its widespread use in research and training programs, the structure of the CTRS has not been examined in a sample of community mental health clinicians with adult and child clients. The current study addressed this gap. The sample consisted of 355 clinicians and 1298 CBT sessions scored using the CTRS. High interrater reliability was observed and factor analysis yielded separate structures for child and adult treatment sessions. These structures were not consistent with previous factor analyses conducted on the scale. Findings suggest that the CTRS is a reliable measure of CBT in community mental health settings but that its structure may depend on the clinical population measured. Additionally, the factor structure can provide guidance for delivering feedback in training and supervision settings.

Keywords Treatment fidelity · Cognitive Therapy Rating Scale · Community mental health · Factor analysis · Psychometrics

Cognitive-Behavioral Therapy (CBT) is the most widely studied form of psychotherapy (Beck and Haigh 2014; Gaudiano 2008; Hofmann et al. 2013) and has a large base of empirical support for treating various mental health disorders (see Hofmann et al. 2012 for a review). Beyond its strong history, CBT also continues to represent innovation in mental health treatment, as it is refined and implemented to serve new clinical populations in diverse settings (Beck and Haigh 2014; Creed et al. 2016a). Measuring and ensuring high quality CBT and other evidence-based practices (EBPs) increases the likelihood that clients will experience the benefits demonstrated in clinical trials. As such, a variety of measures have been developed to assess the quality (i.e., fidelity) of clinicians delivering CBT (Muse and McManus 2013). However, despite the strong emphasis placed on

transporting EBPs to real-world practice, very little is known about how fidelity measures developed for clinical trials perform in these settings.

Treatment Fidelity

Treatment fidelity, or the adherence to and competence with core features of a specific treatment, is a key implementation outcome for a number of reasons (Proctor et al. 2011). For example, greater treatment fidelity is predictive of desirable client outcomes across numerous mental health disorders (Hogue et al. 2008; Schoenwald et al. 2008; Strunk et al. 2010; Stirman et al. 2013a). Strategies that increase fidelity to an implemented EBP increase the likelihood that clients will benefit from that treatment. Indeed, there is a large consensus on the need to verify fidelity for EBPs (Rollins et al. 2016). Yet, clinicians often report making modifications to standardized treatments in routine care (Aarons et al. 2012; Cook et al. 2014; Stirman et al. 2013b). Without valid fidelity measures these modifications go unquantified, which prevents our understanding of how adaptations affect client outcomes. That is, valid fidelity measures allow us to differentiate between "flexibility within fidelity", which is

✉ Torrey A. Creed
  tcreed@pennmedicine.upenn.edu

[1] Aaron T Beck Psychopathology Research Center, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19107, USA

[2] Aaron T. Beck Psychopathology Research Center, University of Pennsylvania Perelman School of Medicine, 3535 Market St., Room 3038, Philadelphia, PA 19104, USA

linked to desirable client outcomes, and poor adherence to a treatment, leading to poor client outcomes (Hamilton et al. 2008; McHugh et al. 2009).

Measurement of fidelity is also particularly important in the training of clinicians. Research suggests that many therapists who believe they are already delivering an EBP in their regular practice may not actually do so with fidelity (Creed et al. 2016b) and a significant proportion of clinicians continue to fail to deliver the therapy with fidelity even after training (Miller et al. 2004; Stirman et al. 2012). Fidelity measurement provides a structure to inform clinicians whether they are successfully delivering the chosen EBP and trainers use this structure to examine areas of skill deficit and provide feedback to bolster those skills. Integration of fidelity measures into the training of therapists is a useful strategy for increasing fidelity through the provision of feedback and monitoring of skill (Sholomskas et al. 2005; Waltman et al. 2018), but only if the psychometric properties of those measures are understood within the treatment population and context.

Cognitive Therapy Rating Scale

Among the over 60 different measures of CBT fidelity that were identified in a recent review (Muse and McManus 2013), the most common and widely used measure of CBT fidelity is the Cognitive Therapy Rating Scale (CTRS; Young and Beck 1980). Although a revised version of the CTRS exists, the Cognitive Therapy Scale Revised (CTS-R), the primary difference between the two scales is the additional CTS-R item (e.g., "eliciting of appropriate emotional expression") that is subsumed under a different CTRS domain, and the splitting of an CTRS item (e.g., "Focusing on key cognitions and behaviors") into two items that separately assess focus on cognitions and behaviors. Indeed, both scales show the same relation with symptom change in the treatment of depression (Kazantzis et al. 2018). The CTRS remains the gold standard for measuring CBT treatment fidelity in clinical trials (e.g., Borkovec et al. 2002; McManus et al. 2010), studies of effective CBT delivery for a range of disorders (Forand et al. 2011; Keen and Freeston 2008), training programs (Creed et al. 2016a; Lewis et al. 2014) and formal certification (e.g., the Academy for Cognitive Therapy).

Previous examination of the CTRS found it to display strong psychometrics in clinical trials of CBT, though there remains a paucity of information about its psychometric characteristics in real-world clinical settings and its applicability to the fidelity of child CBT sessions (Fuggle et al. 2012). The CTRS has demonstrated high internal consistencies, both in the original investigation of its psychometrics (Dobson et al. 1985) and in a recent study (McManus et al.

2010). Scores on the CTRS increased following CBT training sessions (Simons et al. 2010; Westbrook et al. 2008) and differentiated between low and high quality sessions (Vallis et al. 1986). Additionally, studies have found it to provide predictive validity. Trepka et al. (2004) found that CTRS total scores significantly correlated with self-rated depression scores. Similar results were reported for clinician, but not patient, rated scores (Shaw et al. 1999). Importantly, there is mixed evidence on the link between CTRS competence and treatment outcomes (e.g., Branson et al. 2015). The CTRS has demonstrated moderate to high inter-rater reliability intraclass correlations (ICC; Dimidjian et al. 2006; McManus et al. 2010; Westra et al. 2009; Creed et al. 2016a), with few exceptions (e.g., Jacobson and Gortner 2000; Rozek et al. 2018). The lack of uniform standards around training for coding and obtaining reliability across studies may have contributed to the mixed evidence for the reliability of the CTRS. Indeed, the majority of studies reporting inter-rater reliability statistics have reported at least moderate to strong ICC. Yet, even with strong psychometrics, and high utility and use in research and training, there is a dearth of research exploring the factor structure of the CTRS.

The CTRS was originally developed to contain two theorized factors: 'general therapeutic skills' and 'cognitive-behavioral skill' (Young and Beck 1980). Among the empirical examinations of these factors, findings have not uniformly supported this division. For example, the primary study of the CTRS factor structure did find a two-factor solution but specific items did not load on the domains as hypothesized (Vallis et al. 1986). One factor explained 8.9% of the score variance and consisted of 3 items (i.e., Agenda, Pacing, and Homework) that did not match with the expected structure. The other factor accounted for 64.8% of score variance and included the remaining 8 items (i.e., Feedback, Understanding, Interpersonal Effectiveness, Collaboration, Guided Discovery, Focusing on Key Cognitions and Behaviors, Strategy for Change, and Application of CBT Techniques). Other studies have reported separate structures. For example, a three factor structure, measuring 'general interview procedures', 'interpersonal effectiveness', and 'specific CBT techniques,' was examined in one study and found to have significant correlations among them (Trepka et al. 2004). However, this was not a formal factor analysis but a division of the measure into subscales based on a priori decisions. A separate study employed a similar procedure, where a priori decisions were the rationale for dividing the measure into the same three clusters (Westbrook et al. 2008).

In addition to the limitations that arise when using an a priori model to determine factor structure (i.e., the a priori model may not be valid), there are also important considerations regarding the clinician sample in these studies. For example, Vallis et al. (1986) used a small sample of

doctoral level clinicians (i.e., Ph.D. or M.D.) trained as part of a research training program, and Westbrook et al (2008) and Trepka et al. (2004) used samples of clinical psychologists. Although the sample in McManus et al's (2010) was also from a cohort of trainees in a CBT training program, the majority of clinicians were doctoral level practitioners (e.g., psychologists or psychiatrists). The authors did not report the nature of their practice settings. All clinicians in these studies worked primarily with adults. The results of these studies, except for possibly McManus et al (2010)'s, are most pertinent to those clinicians working with adults in university or research settings, with ample supervision, training, and resources. McManus et al (2010)'s sample may reflect a sample of clinicians working in the community, though it is unclear, but no information on the factor structure of the CTRS is provided. Further study is necessary to determine whether the CTRS factor structure would be replicated when therapy is delivered by clinicians working in non-academic practice settings with diverse populations, age groups, and presenting problems. This is particularly true given the sharp increase in effort aimed at implementing CBT in community mental health settings (McHugh and Barlow 2010). Clinicians in community mental health settings may face unique and different challenges related to funding issues, staff turnover, leadership challenges, and low levels of technical support (Fixsen et al. 2005). These clinicians typically also have less training and supervision than study therapists in clinical trials and often treat more diverse populations in regard to presenting problem (i.e., all who present in a community clinic versus those who meet criteria for a clinical trial) and demographics (e.g. child versus adult; Creed et al. 2016a).

The Current Study

As noted earlier, treatment fidelity measures such as the CTRS are an integral part of understanding and measuring EBP implementation efforts (Herschell et al. 2010). They help us understand how to achieve desirable outcomes, enhance efficacy, replicate successful programs, and measure performance over time (Essock et al. 2015). However, if these fidelity measures are only examined within samples of research clinicians working within university settings, they may not represent community clinicians and findings based on them may be misleading. Exploring fidelity measures in samples of community clinicians is needed to assure robust conclusions from their use. That is, factor structures differ for different populations and community mental health clinicians represent a unique population, due to the environment in which they work, when compared to research clinicians working in university settings (Morse et al. 2012). By understanding the factor structure of the CTRS in a sample of

community clinicians, we can provide better training and feedback to community mental health clinicians and more accurately assess implementation efforts of CBT in community mental health settings.

The primary aim of the current study was to examine the factor structure of the CTRS within a sample of outpatient community mental health clinicians who treat both adults and children with diverse presenting problems. Specifically, we sought to examine whether the two-factor structure of the CTRS would be reproduced in this sample. To ensure the factor analysis is valid, we will also examine interrater reliability. Additionally, given that the CTRS factor structure had not previously been examined using child CBT sessions despite extensive research in this field (Fuggle et al. 2012; James et al. 2013), we also sought to examine variability in structure between treatment sessions with children and adults. We hypothesized that the two-factor structure proposed by Vallis et al. (1986) would fit this sample and would be invariant across clinicians working with children or adults. To our knowledge, this is the first investigation of this measure in a broad community mental health population and encompassing diverse populations.

Method

Procedures

Data for the current study were collected as part of an ongoing implementation effort and related program evaluation project, The Beck Community Initiative (BCI). The BCI is a community-academic partnership to provide CBT training and implementation support for community mental health clinicians in under-resourced settings. The BCI has successfully implemented CBT in a wide range of settings and populations (e.g., chronic homelessness psychosis, substance abuse, and child and family services; Creed et al. 2013; Pontoksi et al. 2016; Riggs and Creed 2017). The BCI protocol has been described in detail elsewhere (Creed et al. 2014, 2016a), though a brief summary is presented here for context.

When a community agency partners with the BCI, they receive an intensive 22-h in-person didactic in the theory and practice of CBT followed by 6 months of direct consultation to improve CBT skill. Clinicians receive weekly group consultation that includes feedback on audiotaped sessions from doctoral level experts in CBT. Subsequent to the training period, a consultation leader is identified to help facilitate an ongoing peer consultation group and to assist in the enrollment of additional clinicians. These additional clinicians are provided access to an extensive online training (German et al. 2017) followed by 6 months of peer consultation, including tape review. Throughout the intensive

training phase, tailored strategies target the sustainability of CBT (e.g., adaptation of policies and procedures to support the model, consultation with supervisors), and after intensive training ends, the consultation team continues to meet periodically with the agency to provide support and promote sustainability. Certification in CBT is provided to clinicians who meet competency requirements, based on CTRS total scores. Clinicians attempt recertification 2 years after completion of their initial certification.

Fidelity rating is conducted by doctoral level experts in CBT prior to training, post-workshop, mid-consultation (3 months post workshop), end of consultation (6 months post workshop), and for recertification purposes (2 years after certification) using the CTRS. More information about the scoring process is reported below. Prior to obtaining audio recordings, clients of participating clinicians provide consent for therapy sessions to be audio recorded and reviewed by the BCI instructors. The current study was deemed exempt by the University of Pennsylvania Institutional Review Board.

Participants

Participants were 355 outpatient clinicians enrolled in the BCI. Outpatient clinicians were selected from the larger group of clinicians (e.g., inpatient, residential, intensive outpatient) who participated in the BCI in order to create a homogenous sample across both adult and child clinicians. That is, because all child clinicians were based in outpatient settings, a sample of adult clinicians from outpatient settings was selected in order to facilitate comparisons, rather than including settings in which the full milieu was trained in CBT principles (e.g., Pontoksi et al. 2016; Riggs and Creed 2017). This selection also facilitated a more appropriate comparison with previously published factor analyses. Clinicians were mostly female (81.7%; 17.5% male) and varied in age from 23 to 74 ($M = 36.10$, $SD = 10.48$). The sample of clinicians was 31.3% Caucasian, 14.4% African American, 5.1% Hispanic, 3.9% Asian, 1% Native American and Native Hawaiian or Pacific Islander, 2.3% other, and 42% chose not to provide this information. The highest degree obtained for most clinicians was a Master's degree (85.4%), with 6.8% having obtained a doctorate or MD, and 2.5% as completing some doctoral work. Most clinicians identified their primary role was as a therapist (81.4%), with few social workers (5.9%), psychologists (1.1%), and creative arts therapists (1.1%). Clinicians previous knowledge of CBT was rated as "nothing" (0.2%), "only the basics" (72.4%), and "a great deal" (14.1%). Of the current sample of clinicians, 181 (51%) were child clinicians whereas 174 (49%) were adult clinicians.

From the 355 clinicians, a total of 1298 audio recordings were rated ($n = 585$ for child therapy sessions, $n = 713$ for adult therapy sessions). All audio recordings were active treatment cases. Clinicians submitted at least one therapy session, though clinicians averaged more than three sessions submitted, with a range from one to five sessions. Clinicians chose a CBT session recorded within 2 weeks of the submission due date at each time point; thus specific clients were not followed over time and sessions submitted at different time points represent different clients at different points in their treatment. However, it is possible clinicians submitted sessions from the same client, at two different time points in their treatment. To ensure clinicians were familiar with and conducting CBT in rated audio sessions, the current study used only those audio recordings that were recorded following the completion of in-person or online CBT training (i.e., baseline audio recordings were excluded).

Measures

The Cognitive Therapy Rating Scale (CTRS; Young and Beck 1980; Vallis et al. 1986) was used to assess therapist fidelity to Cognitive-Behavioral Therapy within treatment sessions. The CTRS consists of 11 items (see Table 1) that are rated on a 0 to 6 Likert-type scale. Total scores range from 0 to 66, and a score of 40 is the cutoff for determining competence (Shaw et al. 1999). Doctoral-level CBT experts, either clinical psychologists or postdoctoral fellows in clinical psychology who served as instructors on the BCI, evaluated audio recorded sessions and rated therapist CBT skill on each item, the sum of which were calculated for a total score. There were 24 raters over the course of the study who scored an average of approximately 54 therapy sessions (number of therapy sessions scored by raters ranged from 3 to 311). Initial calibration was achieved by all raters prior to that individual scoring the current sample of CBT audio sessions. Training audio were archived audio recordings from community mental health clinicians who had previously participated in the BCI. Prior to initial calibration, all raters undergoing training were provided the CTRS scoring manual (Young and Beck 1980), as well as training materials developed by the research team with scoring rules to support interrater reliability. During this initial calibration period, raters undergoing training also observed formal calibration meetings held among trained raters to discuss scoring and prevent drift. During calibration, raters were provided feedback on their scores until accurate scores were obtained in 5 consecutive sessions. Accuracy was determined by rating each item scores within one point of a gold-standard score, as well as agreement about whether the total score reflected competence (total ≤ 40).

Additionally, in order to reduce the effect of rater drift on scores, regular calibration meetings were held among raters following their initial reliability training. All raters independently scored sessions blind to the scores of other

Table 1 Means, standard deviations, ICC and item-total correlations of CTRS total score and items by clinical population

	Total sample (N = 1298) M (SD)	Child treatment (n = 585) M (SD)	Adult treatment (n = 713) M (SD)	Group differences t	ICC	Item-total correlation
Agenda	2.18 (1.61)	2.45 (1.65)	1.96 (1.54)	5.63**	.81	0.78**
Feedback	2.14 (1.42)	2.29 (1.40)	2.01 (1.43)	3.44*	0.86	0.76**
Understanding	3.10 (0.81)	3.10 (0.90)	3.08 (0.73)	0.36	0.77	0.68**
Interpersonal effectiveness	3.83 (0.89)	3.89 (0.95)	3.78 (0.83)	2.16*	0.75	0.52**
Collaboration	3.14 (0.99)	3.18 (1.08)	3.10 (0.90)	1.46	0.86	0.74**
Pacing and efficient use of time	2.76 (1.03)	2.85 (1.09)	2.68 (0.97)	2.99*	0.85	0.74**
Guided discovery	2.54 (0.98)	2.65 (1.00)	2.45 (0.94)	3.72**	0.77	0.75**
Focusing on key cognitions or behaviors	2.59 (1.24)	2.70 (1.26)	2.50 (1.22)	2.86*	0.86	0.81**
Strategy for change	2.39 (1.38)	2.57 (1.26)	2.24 (1.33)	4.31**	0.89	0.83**
Application of CBT techniques	2.07 (1.31)	2.23 (1.36)	1.93 (1.25)	4.18**	0.88	0.85**
Homework	1.83 (1.43)	1.77 (1.45)	1.88 (1.42)	1.35	0.94	0.72**
Total score	28.56 (9.85)	29.70 (10.62)	27.63 (9.07)	3.79**	0.89	-

ICC intraclass correlation, *CTRS* Cognitive Therapy Rating Scale, *CBT* cognitive-behavior therapy

* $p < .05$, ** $p < .001$

raters, and then used the meetings to discuss the rationale for individual item scores, rather than group scoring sessions during meetings. A subset of 45 treatment sessions was used to obtain interrater reliability, calculated using intraclass correlations with a one-way random effects model because not all sessions were rated by the same raters. Assumptions for calculating ICC were met, including approximately normally distributed data and homogenous variance. High interrater reliability was obtained for CTRS total scores (ICC = .89). Although some previous studies have shown discrepant interrater reliability (e.g., Rozek et al. 2018), the current study is consistent with a number of studies finding acceptable interrater reliability with doctoral-level raters (e.g., McManus et al. 2010). Individual item ICC are presented in Table 1.

The CTRS items compose two theory-driven subscales: General Therapeutic skills and Cognitive-Behavioral Therapy skill. General Therapeutic skills consist of the items: Agenda, Feedback, Understanding, Interpersonal Effectiveness, Collaboration, and Pacing. Cognitive-Behavioral Therapy skill consists of the items: Guided Discovery, Focusing on Key Cognitions and Behaviors, Strategy for Change, Application of CBT Techniques, and Homework. However, as noted previously, factor analysis has demonstrated that a different two-factor solution may be appropriate (Vallis et al. 1986). Studies examining the scale's psychometrics have found it to be moderately reliable, $r = .59$, a valid measure, and sensitive to changes in the quality of CBT skill (Dobson et al. 1985; Vallis et al. 1986).

Statistical Analyses

Data were analyzed using IBM SPSS Statistics (SPSS) version 21 and AMOS version 21 (Arbuckle 2012). AMOS uses maximum likelihood estimation procedures to determine model parameters. Prior to testing study hypotheses, descriptive statistics, item-total correlations, and group differences were calculated. Demographic and CTRS score differences between child clinicians and adult clinicians were examined using t tests and ANOVA where appropriate.

In order to investigate whether the original factor structure found by Vallis et al. (1986) fit the current sample, a confirmatory factor analysis (CFA) was performed. A CFA was then performed with child clinicians and adult clinicians separately to test for multiple group invariance. To assess model fit, five separate indicators were examined: Chi square (χ^2), Tucker Lewis Index (TLI), root mean square error of approximation (RMSEA), standard root mean square residual (SRMR), and comparative fit index (CFI). Using different fit indices allows for a broad estimation of goodness of fit for the full model, while not relying on any single indicator that may have limitations. Goodness of fit is indicated by a nonsignificant χ^2, TLI > .90, RMSEA < 0.06, SRMR < 0.08, and CFI > 0.95 (Hu and Bentler 1999). If the factor structure was found to be ill-fitting for the current data, post hoc exploratory factor analysis (EFA) would be used to explore underlying structure that would best fit the data. EFA and CFA would be conducted using the same sample to ensure that if differences were observed they were due to methodological explanations, rather than substantive

ones (i.e., the difference in factors cannot be explained differences in samples; Van Prooijen and Van Der Kloot 2001). EFA and CFA, due to differences in the statistical basis for each analysis, provide unique information about the factor structures examined essential to the analysis. Although there are strengths to using separate samples for EFA and CFA, sample size considerations and the benefit of parsimonious model building (Patil et al. 2008) led us to conduct EFA and CFA in the same sample.

As noted above, clinicians had more than one therapy session included in the sample, which may violate the assumption of independence. However, because sessions were obtained at different stages of the training and consultation process, rated independently, and clinicians were allowed to select different clients, the sessions are likely not significantly nested. Additionally, because the focus of the study was on the factor structure of the CTRS at the individual therapy session level, not the clinician level, this nested structure does not necessarily impact the analyses (Huang 2016). Indeed, ICC for this nested structure was small (ICC = .023) indicating sessions are essentially independent (Thomas and Heck 2001).

Results

Descriptive and Correlational Analyses

Means, standard deviations, item-total correlations, and ICC of the CTRS items and total score are presented in Table 1. The CTRS demonstrated high levels of interrater reliability, with ICC ranging from .75 to .94 for individual CTRS items and .89 for the CTRS total score. Additionally, all items were moderately to highly correlated with the total score and the subscale scores (i.e., general therapeutic skill and CBT skill) were highly correlated with each other $r = .82$, $p < .001$. Differences in specific CTRS items were observed between child and adult therapy sessions. Scores for Agenda, Feedback, Interpersonal Effectiveness, Pacing, Guided Discovery, Focusing on Key Cognitions or Behaviors, Strategy for Change, and Application of CBT techniques were all

found to significantly differ between child and adult sessions. Importantly, effect sizes for these differences were uniformly in the small range, with Cohen's d values ranging from 0.16 (Focusing on Key Cognitions or Behaviors) to 0.31 (Agenda).

Factor Structure of the CTRS

The fitness of the factor structure proposed by Vallis et al. (1986) was examined using confirmatory factor analysis (CFA) with the full sample. When taken together, the model fit indices for this model (see Table 2, Model 1) were not acceptable based on the criteria listed above (Hu and Bentler 1999). As such, we did not conduct a CFA to examine invariance of this model based on child or adult therapy sessions. However, post hoc analyses were conducted to examine the factor structure in the current sample. An exploratory factor analysis (EFA) with oblique rotation (Costello and Osborne 2005) was performed to explore the underlying factor structure of the data. Additionally, a Kaiser–Meyer–Olkin (KMO) test was conducted to ensure the sample was appropriate for conducting factor analysis. The KMO value was 0.92, greater than the 0.70 cutoff, indicating the items are suitably factorable (Beavers et al. 2013). The results from the EFA (see Table 3) show that no item had a factor loading below .30 and thus, no items were dropped from the analyses. Though two factors were extracted from the items, explaining 59.3% of the variance, the loadings differed from the Vallis et al. (1986) model.

A CFA was performed using a structure based on the results of the EFA. In this model, errors between homework and agenda, strategy for change and application of CBT technique, and feedback and agenda were correlated given partial overlap in scoring of these items (e.g., receiving client feedback is an important part of agenda setting; Landis et al. 2009). Model fit indices (see Table 2, Model 2) indicated that this model had adequate to good model fit. Although the Chi square (χ^2) test remains significant, it not a preferred measure of fit due to insensitivity when used in large samples (Byrne 2004; Hu and Bentler 1999). RMSEA, SRMR, and CFI are all less sensitive to sample

Table 2 Summary of confirmatory factor analysis model fit indices

Model	Chi square (χ^2)	TLI	RMSEA	SRMR	CFI
1. Vallis et al. two factor replication	1283.99	0.82	0.15	0.07	0.86
2. Two factor model based on EFA	387.42	0.94	0.08	0.04	0.96
3. Configural invariance unconstrained model	482.62	0.95	0.06	0.03	0.95
4. Metric invariance constrained model	524.30	0.94	0.06	0.05	0.95
5. Adult treatment session model	224.09	0.94	0.08	0.05	0.96
6. Child treatment session model	219.20	0.95	0.09	0.04	0.96

TLI Tucker Lewis Index, *RMSEA* root mean square error of approximation, *SRMR* standard root mean square residual, *CFI* comparative fit index, *EFA* exploratory factor analysis

Table 3 Results of exploratory factor analyses

Item	Whole sample		Adult treatment sessions		Child treatment sessions
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1[a]
Agenda	**.82**	.09	**.78**	.08	**.73**
Feedback	**.81**	.09	**.81**	.13	**.73**
Understanding	.12	**.75**	.14	**.72**	**.75**
Interpersonal effectiveness	.07	**.75**	.11	**.70**	**.62**
Collaboration	.36	**.49**	.37	**.47**	**.78**
Pacing and efficient use of time	**.50**	.30	**.48**	.36	**.72**
Guided discovery	**.48**	.35	**.48**	.36	**.76**
Focusing on key cognitions or behaviors	**.61**	.26	**.66**	.17	**.84**
Strategy for change	**.74**	.14	**.75**	.10	**.84**
Application of CBT techniques	**.73**	.18	**.78**	.13	**.84**
Homework	**.77**	.11	**.74**	.11	**.68**
Eigenvalue	6.17	1.17	5.66	1.39	6.30
% Variance explained	52.52	6.77	47.55	8.27	57.28

Factor loadings > .40 are in boldface

[a]1 Factor solution

size and were used to evaluate the model. Given the good model fit obtained with this model, a multi-group analysis was performed to examine the structural equivalence of the model across child and adult treatment sessions.

The test for configural invariance across child and adult treatment sessions was conducted in accordance with Byrne's (2004) recommendations. The freely-estimated, unconstrained model, analyzed across the two groups, yielded good model fit (see Table 2, Model 3). Additionally, metric invariance was tested across child and adult treatment sessions and yielded good model fit (see Table 2, Model 4). However, the Chi square (χ^2) difference test between these models resulted in a significant Chi square, $\chi^2 (12) = 41.68$, $p < .001$, indicating significant variability across groups. Post hoc analyses were conducted to examine group differences.

In order to explore the underlying factor structure within each group (Matsunaga 2010) exploratory factor analyses (EFA) with oblique rotation were conducted separately for child and adult treatment session groups. Kaiser–Meyer–Olkin (KMO) values were 0.90 for adult treatment sessions and 0.93 for child treatment sessions, greater than the 0.70 cutoff, indicating the items are suitably factorable (Beavers et al. 2013). The results of these EFAs are presented in Table 3 and show no item with a factor loading below .30; thus, all items were retained for each group. However, the adult treatment group EFA yielded a two-factor structure similar to that obtained for the whole sample (Table 2, Model 2), whereas the child treatment group EFA yielded a one-factor structure. Confirmatory factor analysis (CFA) was then conducted on each group separately. Results of these CFAs are found in Table 2 (Models 5 and 6).

Factor loadings and parameter weights for each the child treatment sessions and the adult treatment sessions are shown in Figs. 1 and 2. In both models, similar to Model 2, errors between homework and agenda, strategy for change and application of CBT technique, and feedback and agenda were correlated. Additionally, for the child treatment group, the errors between understanding and interpersonal effectiveness were also correlated. For the adult treatment group, when accounting for correlated errors, pacing and efficient use of time was found to load on the general therapy skill factor. This differed from the EFA. Both models demonstrated adequate to good model fit, when model fit indices were examined holistically. Although RMSEA is greater than the suggested cutoff, it has been found to be especially conservative at smaller sample sizes (Hu and Bentler 1999) and may require larger sample sizes in each individual group to meet suggested cutoffs.

Discussion

The current study examined the factor structure and reliability of the CTRS in a sample of child and adult clinicians working in outpatient community mental health settings. We hypothesized that the two-factor structure proposed by Vallis et al. (1986) would best fit the current sample and that this factor structure would be invariant across clinicians who work with children or adults. Our data did not support these hypotheses. Instead, our results demonstrated a unique two-factor solution for adults and a one-factor solution for

Confirmatory factor analysis of the CTRS for child treatment sessions
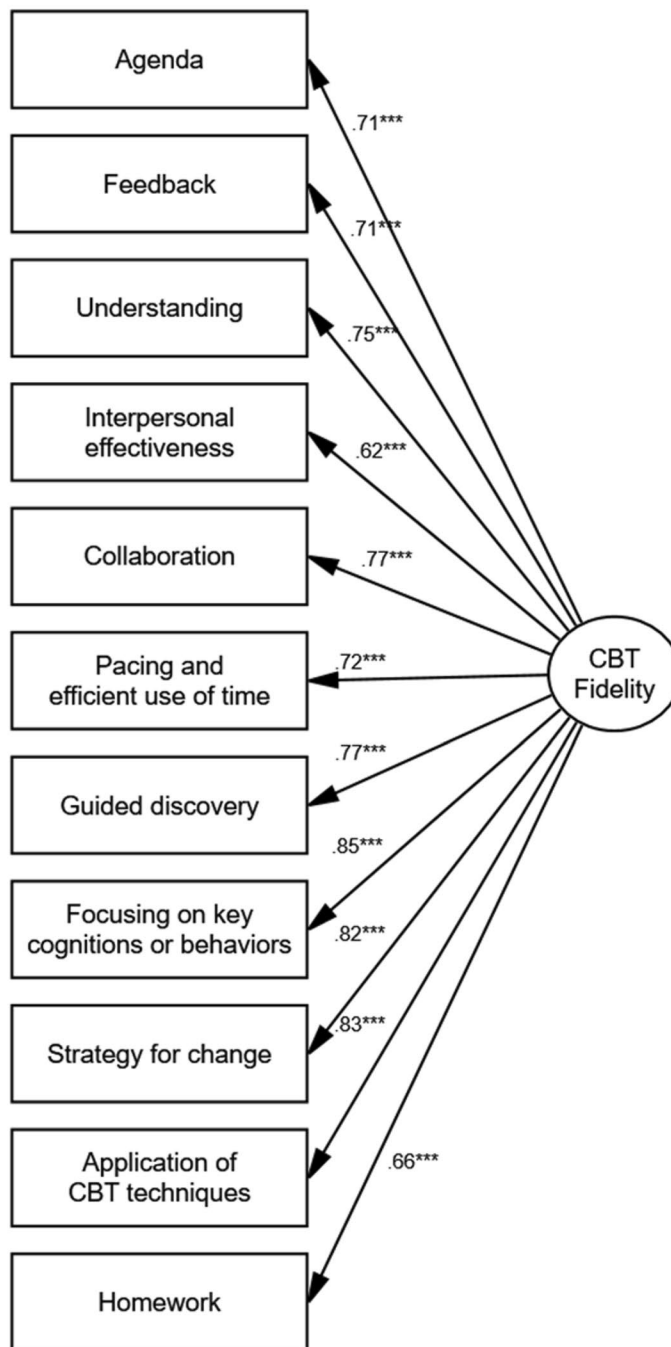
Note. CBT = Cognitive-behavior therapy

*** *p* < .001

children that suggests that the CTRS structure may differ based on clinical populations and settings.

Overall, the CTRS was found to be a reliable measure of cognitive-behavior therapy for children and adults in community mental health settings. Although previous research has been inconsistent on whether the scale demonstrates high levels of inter-rater reliability (e.g., Jacobson and Gortner 2000), this study was consistent with other studies

Confirmatory factor analysis of the CTRS for adult treatment sessions
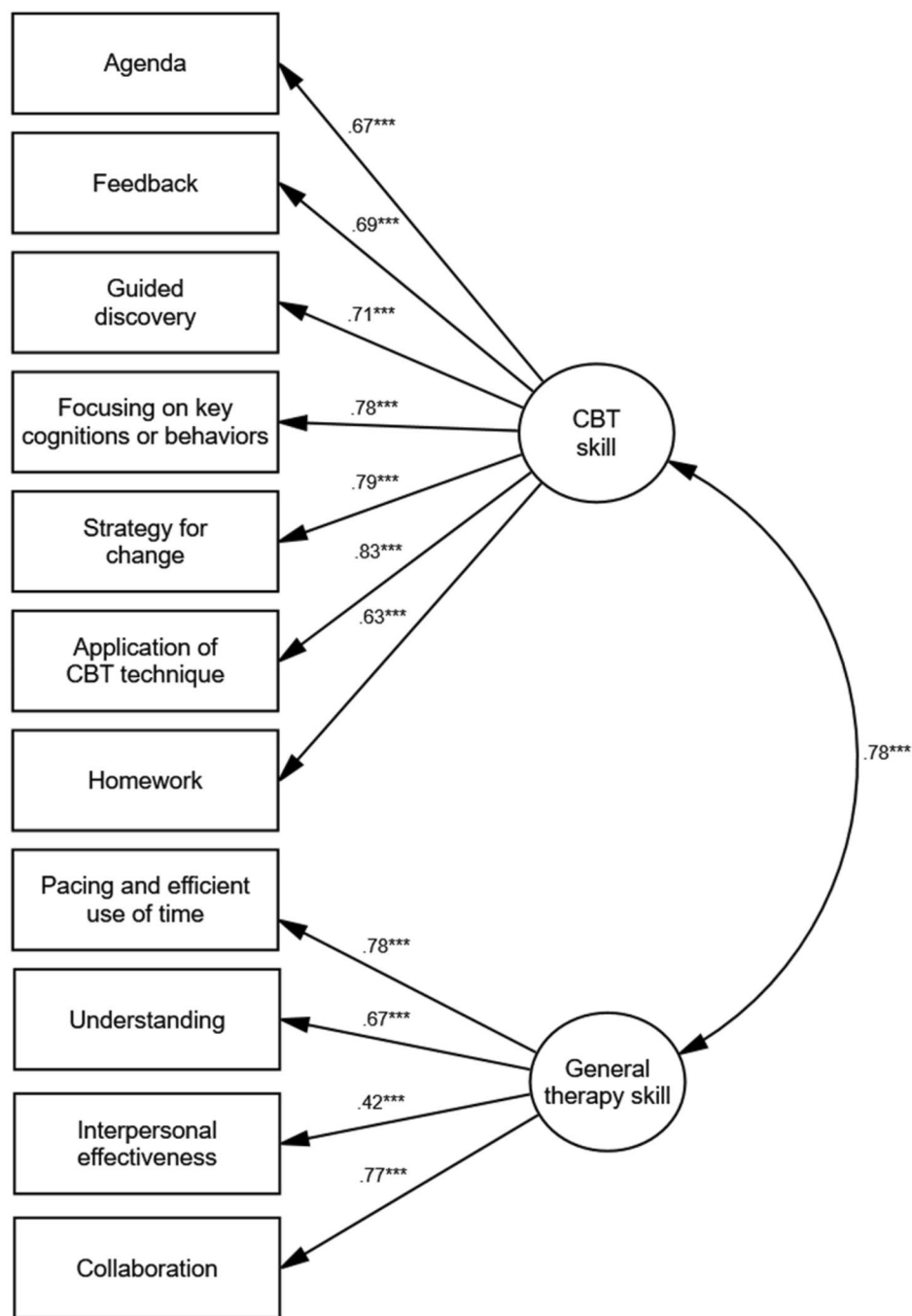
Note. CBT = Cognitive-behavior therapy

*** *p* < .001

that showed high inter-rater reliability when the raters were trained CBT experts (e.g., McManus et al. 2010). Indeed, some studies that have shown lower inter-rater reliability have used non-CBT expert or undergraduate level research assistants (Rozek et al. 2018). High inter-rater reliability was shown not only for total scores, which are used for certification purposes, but for individual item scores as well, which may be integral for training purposes. That is, agreement

across both total and item scores suggests not only are competence scores reliably obtained, but also expert trainers can reliably identify specific strengths and weaknesses from individual items on the CTRS. Trainers can use this information to formulate individualized feedback to strengthen areas of difficulty. Importantly, in keeping with best practices, rigorous reliability training ensured that all raters were calibrated before rating any study tapes, and ongoing group calibration meetings were used to prevent drift.

As noted above, the best fitting model for the entire sample was a two-factor structure, though the specific items differed from those previously found to load on the two factors. In the current two-factor solution the general therapeutic skill factor consisted of: understanding, interpersonal effectiveness, and collaboration. The Cognitive-Behavioral Therapy skill factor consisted of: agenda, feedback, pacing and efficient use of time, guided discovery, focusing on key cognitions or behaviors, strategy for change, application of CBT techniques, and homework. This solution explained lower amounts of variance than Vallis et al.'s (1986) model, though it did achieve good model fit. Given that no other studies have conducted factor analysis on the scale, it is difficult to parse whether differences in the factor structure are due to differences in sample characteristics (e.g., community mental health clinicians versus research-trained clinicians) or variability inherent to the CTRS. However, given that differences do exist among the populations observed (Fixsen et al. 2005; Proctor et al. 2009), it is plausible that the structure observed is due to unique aspects of community mental health clinicians, such as low levels of support and supervision, and treatment of diverse presenting problems (Garety et al. 2017). Clinicians in the current sample may have benefited from the support of the BCI training program and may not extend to those clinicians who work in community mental health agencies without access to supplemental trainings.

Further examination of the factor structure did not support invariance between child and adult treatment sessions. That is, two different factor structures were observed to be the best fitting models: one for therapy sessions with children and the other for therapy sessions with adults. For adult treatment sessions, a two-factor solution was appropriate. The general therapeutic skills factor consisted of the items understanding, interpersonal effectiveness, collaboration, and pacing and efficient use of time. The Cognitive-Behavioral Therapy skill factor consisted of agenda, feedback, pacing and efficient use of time, guided discovery, focusing on key cognitions or behaviors, strategy for change, application of CBT techniques, and homework. This separation of items is in line with the theoretical differentiation in CTRS measurement. In other words, the CBT skill factor consists of those items relevant to performing CBT specific capabilities well (e.g., session structure and interventions). The

general therapy skill factor consists of those items which are not unique to CBT, and may be demonstrated in non-CBT sessions.

Interestingly, pacing and efficient use of time loaded highly on both factors in our exploratory factor analysis (EFA) but was found to load only on the general therapy skill factor in our confirmatory factor analysis (CFA). Conceptually, high loadings on both factors is unsurprising given that pacing and proper session structure is necessary for quality CBT interventions to be performed but must not interfere with collaborative and interpersonal aspects of treatment. This has been supported by research that has shown the use of CBT session structure to correlate with treatment response (e.g., Ginsburg et al. 2012). Additionally, this is similar to the findings of Vallis and colleagues (1986). In their study 'pacing' had high factor loadings (> .5) on both factors as well. Importantly, under the constraints of the CFA, this item was found to have a low loading on CBT skill, which may suggest that pacing and efficient use of time is more related to general therapeutic skill.

For child treatment sessions, a one-factor model best represented the data. This result, not previously demonstrated in the literature, suggests that the CTRS structure may not be consistent across all clinical populations and settings. It also suggests that a common skill may underlie all items. Perhaps child clinicians with lower general therapeutic skills (e.g., collaboration) have difficulty applying CBT techniques, whereas those with higher general skills are able to apply CBT techniques more readily. As these skills tend to appear together, those who train and supervise clinicians working with children may need to approach training more broadly, teaching a more integrated skillset. Though research on this is sparse, this hypothesis is consistent with the finding that therapist flexibility and collaboration is related to child engagement (Chu and Kendall 2009; Hamilton et al. 2008) and that child engagement is related to treatment response (Chavira et al. 2014). Further, this finding suggests the need for future research to examine the CTRS in different clinical settings with different clinical populations in order to best understand its structure.

It has been suggested elsewhere that competencies required for CBT in children may be distinct from those required for CBT in adults (Roth et al. 2011). Indeed, Stallard (2005) describes the importance of partnering with parents/caregivers, matching the intervention to child developmental level, presenting CBT information creatively and flexibly, and engagement, among others, as specific competencies needed for effective CBT treatment for children. This was the basis for the creation of a separate scale specifically aimed at measuring CBT competence for practitioners working with children (i.e., The Cognitive Behaviour Therapy Scale for Children and Young People; Stallard et al. 2014). The scale was created to include the concepts from the

CTRS with specific applicability to children. Although the CTRS does capture some of the domains listed above (e.g., collaboration and interpersonal effectiveness), the current data suggest that CBT competence with children requires broader skills, consistent with Stallard's (2005) theory. That is, the general therapeutic skills domain is important, but separate from CBT skills, for CBT competence with adults; whereas, CBT competence with children does not differentiate between the two. Further research is required to confirm this hypothesis.

These findings are also relevant to CBT implementation efforts. Given the importance of treatment fidelity to the implementation of CBT (e.g., Waltman et al. 2017; Stirman et al. 2013a), valid measurement of treatment fidelity is essential. Although the CTRS is used widely in both research and training settings (Forand et al. 2011), our results are the first exploration of its structure in a sample of community clinicians. This is an essential step in validating the use of the CTRS among this population. Valid fidelity measures serve many stakeholder groups in implementation and training settings, including researchers, trainers, and supervisors (Essock et al. 2015), all of which are informed by the current findings. For example, fidelity research trials (e.g., Stirman et al. 2018) should be cautious regarding the factor structure of the CTRS dependent on setting and population of enrolled clinicians.

In training and supervisory contexts, the feasibility of using the CTRS is a concern. The CTRS is a time and resource intensive measure of fidelity that may not be feasible in many practice settings. However, the results provide a framework for training and delivering feedback efficiently on cases. For example, adult clinicians may be able to independently learn CBT skills and general therapeutic skills effectively; whereas, child clinicians may require a broader more integrated approach. This allows trainers and supervisors to create more effective and efficient training tools and programs, to better observe the successes and failures of clinicians and broader implementation initiatives, and to measure performance accurately to ensure program efficiency and desirable outcomes. This may assist in decreasing the burden of using an observation-based measure of fidelity in community mental health settings.

The findings of the current study should be tempered by certain limitations. The sample of clinicians was obtained from a single implementation program in one urban community mental health care system (Beck Community Initiative; Creed et al. 2014, 2016a), and may not generalize to other settings. Although the project has been shown effective in a range of populations (e.g., Creed et al. 2013; Pontoski et al. 2016), the structure observed may be influenced by the specific training these clinicians have received. That is, more studies in community mental health settings are needed to confirm this structure extends beyond the current sample of

clinicians trained by the BCI. Conducting factor analyses in separate samples will help confirm or disconfirm the stability and robustness of the findings. Additionally, we did not track patient level data in this cohort and thus cannot speak to presenting problems of the clients. This could have an effect on the findings of the study and future studies should include patient level data to determine its effect on the factor structure. Similarly, because clinicians were able to choose which CBT sessions to submit from among all clients with whom they were practicing CBT, we are unable to control for the phase of therapy clients may have been in for that session, whether multiple recordings from the same client were submitted, or whether a session was a more favorable representation of their work. Instead, early, middle, and later stages of the clinicians' training and learning process were represented in the data. Finally, it is important to note that this new factor structure does not necessarily link to improved outcomes for clients. It will be important to examine whether this new factor structure can inform on client outcomes in future research.

Despite these limitations, the current study is an important addition the literature. It is the first study to examine the reliability and factor structure of the CTRS in a sample of community mental health clinicians. Additionally, it is the first study to examine differences in the factor structure of the CTRS based on clinical population. The results were consistent with previous studies showing the CTRS to have high levels of interrater reliability (McManus et al. 2010; Westra et al. 2009). However, factor analyses showed differences in factor structure from previous studies (Vallis et al. 1986) and these differences varied between child and adult therapy sessions. This has important implications for assessing fidelity in community mental health settings whether that occurs within training programs, implementation, or regular supervision. We recommend that feedback and scores be provided within factor structures to ensure that clinicians perform CBT with high levels of fidelity (Waltman et al. 2017). Additionally, future studies are needed to determine whether CBT conducted in other clinical populations and mental health settings is related to different CTRS factor structures from those observed here.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Research Involving Animal Rights This article does not contain any studies with animals performed by any of the authors.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

Aarons, G. A., Miller, E. A., Green, A. E., Perrott, J. A., & Bradway, R. (2012). Adaptation happens: A qualitative case study of implementation of the incredible years evidence-based parent training programme in a residential substance abuse treatment programme. *Journal of Children's Services, 7*, 233–245.

Arbuckle, J. L. (2012). *Amos (Version 21.0) [Computer Program]*. Chicago: IBM SPSS.

Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research & Evaluation, 18*, 1–13.

Beck, A. T., & Haigh, E. A. (2014). Advances in cognitive theory and therapy: The generic cognitive model. *Annual Review of Clinical Psychology, 10*, 1–24.

Borkovec, T. D., Newman, M. G., Pincus, A. L., & Lytle, R. (2002). A component analysis of cognitive-behavioral therapy for generalized anxiety disorder and the role of interpersonal problems. *Journal of Consulting and Clinical Psychology, 70*, 288–298. https://doi.org/10.1037/0022-006X.70.2.288.

Branson, A., Shafran, R., & Myles, P. (2015). Investigating the relationship between competence and patient outcome with CBT. *Behaviour research and therapy, 68*, 19–26.

Byrne, B. M. (2004). Testing for multigroup invariance using AMOS graphics: A road less traveled. *Structural Equation Modeling, 11*, 272–300.

Chavira, D. A., Golinelli, D., Sherbourne, C., Stein, M. B., Sullivan, G., Bystritsky, A., … Bumgardner, K. (2014). Treatment engagement and response to CBT among Latinos with anxiety disorders in primary care. *Journal of Consulting and Clinical Psychology, 82*, 392–403.

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.

Cook, J. M., Dinnen, S., Thompson, R., Simiola, V., & Schnurr, P. P. (2014). Changes in implementation of two evidence-based psychotherapies for PTSD in VA residential treatment programs: A national investigation. *Journal of Traumatic Stress, 27*, 137–143.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*, 1–9.

Creed, T. A., Frankel, S. A., German, R. E., Green, K. L., Jager-Hyman, S., Taylor, K. P., … Williston, M. A. (2016a). Implementation of transdiagnostic cognitive therapy in community behavioral health: The Beck Community Initiative. *Journal of Consulting and Clinical Psychology, 84*, 1116–1126.

Creed, T. A., Jager-Hyman, S., Pontoski, K., Feinberg, B., Rosenberg, Z., Evans, A. C., Hurford, M. O., & Beck, A. T. (2013). The Beck Initiative: A strength-based approach to training school-based mental health staff in cognitive therapy. *International Journal of Emotional Education, 5*, 49–66.

Creed, T. A., Stirman, S. W., Evans, A. C., & Beck, A. T. (2014). A model for implementation of cognitive therapy in community mental health: The Beck Initiative. *Behavior Therapist, 37*, 56–64.

Creed, T. A., Wolk, C. B., Feinberg, B., Evans, A. C., & Beck, A. T. (2016b). Beyond the label: Relationship between community therapists' self-report of a cognitive behavioral therapy orientation and observed skills. *Administration and Policy in Mental Health and Mental Health Services Research, 43*, 36–43. https://doi.org/10.1007/s10488-014-0618-5.

Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., … Atkins, D. C. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology, 74*, 658–670.

Dobson, K. S., Shaw, B. F., & Vallis, T. M. (1985). Reliability of a measure of the quality of cognitive therapy. *British Journal of Clinical Psychology, 24*, 295–300.

Essock, S. M., Nossel, I. R., McNamara, K., Bennett, M. E., Buchanan, R. W., Kreyenbuhl, J. A., … Dixon, L. B. (2015). Practical monitoring of treatment fidelity: examples from a team-based intervention for people with early psychosis. *Psychiatric Services, 66*, 674–676.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network.

Forand, N. R., Evans, S., Haglin, D., & Fishman, B. (2011). Cognitive behavioral therapy in practice: Treatment delivered by trainees at an outpatient clinic is clinically effective. *Behavior Therapy, 42*, 612–623.

Fuggle, P., Dunsmuir, S., & Curry, V. (2012). *CBT with children, young people and families*. London: Sage.

Garety, P. A., Craig, T. K., Iredale, C. H., Basit, N., Fornells-Ambrojo, M., Halkoree, R., … Zala, D. (2017). Training the frontline workforce to deliver evidence-based therapy to people with psychosis: Challenges in the GOALS study. *Psychiatric Services, 69*, 9–11.

Gaudiano, B. A. (2008). Cognitive-behavioral therapies: Achievements and challenges. *Evidence Based Mental Health, 11*, 5–7. https://doi.org/10.1136/ebmh.11.1.5.

German, R. E., Adler, A., Frankel, S. A., Stirman, S. W., Pinedo, P., Evans, A. C., … Creed, T. A. (2017). Testing a web-based, trained-peer model to build capacity for evidence-based practices in community mental health systems. *Psychiatric Services, 69*, 286–292.

Ginsburg, G. S., Becker, K. D., Drazdowski, T. K., & Tein, J. Y. (2012). Treating anxiety disorders in inner city schools: Results from a pilot randomized controlled trial comparing CBT and usual care. *Child & Youth Care Forum, 41*, 1–19.

Hamilton, J. D., Kendall, P. C., Gosch, E., Furr, J. M., & Sood, E. (2008). Flexibility within fidelity. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*, 987–993.

Herschell, A. D., Kolko, D. J., Baumann, B. L., & Davis, A. C. (2010). The role of therapist training in the implementation of psychosocial treatments: A review and critique with recommendations. *Clinical Psychology Review, 30*, 448–466.

Hofmann, S. G., Asmundson, G. J., & Beck, A. T. (2013). The science of cognitive therapy. *Behavior Therapy, 44*, 199–212.

Hofmann, S. G., Asnaani, A., Vonk, I. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research, 36*, 427–440.

Hogue, A., Henderson, C. E., Dauber, S., Barajas, P. C., Fried, A., & Liddle, H. A. (2008). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology, 76*, 544–555.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1–55.

Huang, F. L. (2016). Alternatives to multilevel modeling for the analysis of clustered data. *The Journal of Experimental Education, 84*, 175–196.

Jacobson, N. S., & Gortner, E. T. (2000). Can depression be de-medicalized in the 21st century: Scientific revolutions, counter-revolutions and the magnetic field of normal science. *Behaviour Research and Therapy, 38*, 103–117.

James, A. C., James, G., Cowdrey, F. A., Soler, A., & Choke, A. (2013). Cognitive behavioural therapy for anxiety disorders in children and adolescents. *Cochrane Database of Systematic Reviews, 2013*(6), CD004790. https://doi.org/10.1002/14651858.CD004 690.pub4.

Kazantzis, N., Clayton, X., Cronin, T. J., Farchione, D., Limburg, K., & Dobson, K. S. (2018). The Cognitive Therapy Scale and Cognitive Therapy Scale-Revised as measures of therapist competence in cognitive behavior therapy for depression: Relations with short and long term outcome. *Cognitive Therapy and Research, 42*, 385–397.

Keen, A. J., & Freeston, M. H. (2008). Assessing competence in cognitive-behavioural therapy. *The British Journal of Psychiatry, 193*, 60–64.

Landis, R., Edwards, B. D., & Cortina, J. (2009). Correlated residuals among items in the estimation of measurement models. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity, and fable in the organizational and social sciences* (pp. 195–214). New York, NY: Routledge.

Lewis, C. C., Scott, K. E., & Hendricks, K. E. (2014). A model and guide for evaluating supervision outcomes in cognitive–behavioral therapy-focused training programs. *Training and Education in Professional Psychology, 8*, 165–173.

Matsunaga, M. (2010). How to factor-analyze your data right: Do's, don'ts, and how-to's. *International Journal of Psychological Research, 3*, 97–110.

McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist, 65*, 73–84.

McHugh, R. K., Murray, H. W., & Barlow, D. H. (2009). Balancing fidelity and adaptation in the dissemination of empirically-supported treatments: The promise of transdiagnostic interventions. *Behaviour Research and Therapy, 47*, 946–953.

McManus, F., Westbrook, D., Vazquez-Montes, M., Fennell, M., & Kennerley, H. (2010). An evaluation of the effectiveness of diploma-level training in cognitive behaviour therapy. *Behaviour Research and Therapy, 48*, 1123–1132.

Miller, W. R., Yahne, C. E., Moyers, T. B., Martinez, J., & Pirritano, M. (2004). A randomized trial of methods to help clinicians learn motivational interviewing. *Journal of consulting and Clinical Psychology, 72*, 1050–1062.

Morse, G., Salyers, M. P., Rollins, A. L., Monroe-DeVita, M., & Pfahler, C. (2012). Burnout in mental health services: A review of the problem and its remediation. *Administration and Policy in Mental Health and Mental Health Services Research, 39*, 341–352.

Muse, K., & McManus, F. (2013). A systematic review of methods for assessing competence in cognitive–behavioural therapy. *Clinical Psychology Review, 33*, 484–499.

Patil, V. H., Singh, S. N., Mishra, S., & Donavan, D. T. (2008). Efficient theory development and factor retention criteria: Abandon the 'eigenvalue greater than one' criterion. *Journal of Business Research, 61*, 162–170.

Pontoski, K., Jager-Hyman, S., Cunningham, A., Sposato, R., Schultz, L., Evans, A. C., Beck, A. T., & Creed, T. A. (2016). Using a Cognitive Behavioral framework to train staff serving individuals who experience chronic homelessness. *Journal of Community Psychology, 44*, 674–680.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., … Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges,

and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38*, 65–76.

Proctor, E. K., Landsverk, J., Aarons, G., Chambers, D., Glisson, C., & Mittman, B. (2009). Implementation research in mental health services: An emerging science with conceptual, methodological, and training challenges. *Administration and Policy in Mental Health and Mental Health Services Research, 36*, 24–34.

Riggs, S. E., & Creed, T. A. (2017). A model to transform psychosis milieu treatment using CBT-informed interventions. *Cognitive and Behavioral Practice, 24*, 353–362.

Rollins, A. L., McGrew, J. H., Kukla, M., McGuire, A. B., Flanagan, M. E., Hunt, M. G., … Salyers, M. P. (2016). Comparison of assertive community treatment fidelity assessment methods: Reliability and validity. *Administration and Policy in Mental Health and Mental Health Services Research, 43*, 157–167.

Roth, A. D., Calder, F., & Pilling, S. (2011). *A competence framework for child and adolescent mental health services*. Edinburgh: NHS Education for Scotland.

Rozek, D. C., Serrano, J. L., Marriott, B. R., Scott, K. S., Hickman, L. B., Brothers, B. M., … Simons, A. D. (2018). Cognitive behavioural therapy competency: Pilot data from a comparison of multiple perspectives. *Behavioural and Cognitive Psychotherapy, 46*, 244–250.

Schoenwald, S. K., Carter, R. E., Chapman, J. E., & Sheidow, A. J. (2008). Therapist adherence and organizational effects on change in youth behavior problems one year after multisystemic therapy. *Administration and Policy in Mental Health and Mental Health Services Research, 35*, 379–394.

Shaw, B. F., Elkin, I., Yamaguchi, J., Olmsted, M., Vallis, T. M., Dobson, K. S., … Imber, S. D. (1999). Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *Journal of Consulting and Clinical Psychology, 67*, 837–846.

Sholomskas, D. E., Syracuse-Siewert, G., Rounsaville, B. J., Ball, S. A., Nuro, K. F., & Carroll, K. M. (2005). We don't train in vain: a dissemination trial of three strategies of training clinicians in cognitive-behavioral therapy. *Journal of Consulting and Clinical Psychology, 73*, 106–115.

Simons, A. D., Padesky, C. A., Montemarano, J., Lewis, C. C., Murakami, J., Lamb, K., … Beck, A. T. (2010). Training and dissemination of cognitive behavior therapy for depression in adults: A preliminary examination of therapist competence and client outcomes. *Journal of Consulting and Clinical Psychology, 78*, 751–756.

Stallard, P. (2005). *A clinicians guide to think good feel good: Using CBT with children and young people*. Winchester: Wiley.

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*, 269–282.

Stirman, S. W., Calloway, A., Toder, K., Miller, C. J., DeVito, A. K., Meisel, S. N., … Crits-Christoph, P. (2013a). Community mental health provider modifications to cognitive therapy: Implications for sustainability. *Psychiatric Services, 64*, 1056–1059.

Stirman, S. W., Kimberly, J., Cook, N., Calloway, A., Castro, F., & Charns, M. (2012). The sustainability of new programs and innovations: A review of the empirical literature and recommendations for future research. *Implementation Science, 7*, 1–19.

Stirman, S. W., Marques, L., Creed, T. A., Gutner, C. A., DeRubeis, R., Barnett, P. G., … Jo, B. (2018). Leveraging routine clinical materials and mobile technology to assess CBT fidelity: The Innovative Methods to Assess Psychotherapy Practices (imAPP) study. *Implementation Science, 13*, 1–11.

Stirman, S. W., Miller, C. J., Toder, K., & Calloway, A. (2013b). Development of a framework and coding system for modifications and adaptations of evidence-based interventions. *Implementation Science, 8*, 1–12.

Strunk, D. R., Brotman, M. A., & DeRubeis, R. J. (2010). The process of change in cognitive therapy for depression: Predictors of early inter-session symptom gains. *Behaviour Research and Therapy, 48*, 599–606.

Thomas, S. L., & Heck, R. H. (2001). Analysis of large-scale secondary data in higher education research: Potential perils associated with complex sampling designs. *Research in Higher Education, 42*, 517–540.

Trepka, C., Rees, A., Shapiro, D. A., Hardy, G. E., & Barkham, M. (2004). Therapist competence and outcome of cognitive therapy for depression. *Cognitive Therapy and Research, 28*, 143–157.

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.

Van Prooijen, J. W., & Van Der Kloot, W. A. (2001). Confirmatory analysis of exploratively obtained factor structures. *Educational and Psychological Measurement, 61*(5), 777–792.

Waltman, S. H., Hall, B., McFarr, L., Beck, A. T., & Creed, T. A. (2017). In-session stuck points and pitfalls of community clinicians learning CBT: A qualitative investigation. *Cognitive and Behavioral Practice, 24*, 256–267. https://doi.org/10.1016/j.cbpra.2016.04.002.

Waltman, S. H., Hall, B. C., McFarr, L. M., & Creed, T. A. (2018). Clinical case consultation and experiential learning in cognitive behavioral therapy implementation: Brief qualitative investigation. *Journal of Cognitive Psychotherapy, 32*, 112–127.

Westbrook, D., Sedgwick-Taylor, A., Bennett-Levy, J., Butler, G., & McManus, F. (2008). A pilot evaluation of a brief CBT training course: Impact on trainees' satisfaction, clinical skills and patient outcomes. *Behavioural and Cognitive Psychotherapy, 36*, 569–579.

Westra, H. A., Arkowitz, H., & Dozois, D. J. (2009). Adding a motivational interviewing pretreatment to cognitive behavioral therapy for generalized anxiety disorder: A preliminary randomized controlled trial. *Journal of Anxiety Disorders, 23*, 1106–1117.

Young, J., & Beck, A. (1980). *Cognitive therapy scale: Rating manual. Unpublished manuscript*. Philadelphia: Center for Cognitive Therapy.