

REVIEWS OF ASSESSMENT TOOLS AND DELIVERY

Behavioural couple therapy: measuring therapist competence

I. Rudolf von Rohr^{1,*}, S. Corrie², M.S. Fischer³, D.H. Baucom⁴, M. Worrell² and H. Pote¹

¹Royal Holloway, University of London, Department of Clinical Psychology, Egham TW20 0EX, UK, ²Central and North West London NHS Foundation Trust, London, UK, ³Universitätsklinikum Heidelberg, Germany and ⁴University of North Carolina, USA

*Corresponding author. Email: isabelle.vonrohr@gmail.com

(Received 11 February 2020; revised 18 May 2020; accepted 19 May 2020)

Abstract

This paper outlines a pilot validation study of the newly developed Behavioural Couple Therapy Scale for Depression (BCTS-D). The BCTS-D aims to assess therapist competence in delivering behavioural couple therapy (BCT) and provide therapists with summative and formative feedback on their performance. Completed by both therapist and supervisor, this will aid therapists' reflection on practice and improve performance. This paper will report on two stages in the development of the BCTS-D: (a) a study evaluating content validity, face validity and usability and (b) a focus group examining usability and utility. Both parts of the study were conducted in the context of a BCT training course and included 20 participants who were either BCT supervisors or BCT trainees. Results suggest that the BCTS-D has good face validity, content validity and usability, and provides a useful tool for promoting self-reflection and providing formative feedback. The studies also provided insight into the strengths of the scale and into areas of refinement, and a number of modifications were undertaken to improve the BCTS-D in response to feedback collected. Future research will need to focus on evaluating the psychometric properties of the BCTS-D and continue to adapt the scale to its users' needs.

Key learning aims

- (1) Readers will understand the importance of measuring therapists' competence to improve practice.
- (2) Readers will understand the development of the BCTS-D scale and its initial psychometric properties.
- (3) Readers will know how to use the BCTS-D in everyday clinical practice.
- (4) Readers will know about the challenges of developing a therapist competence measure within a real-world clinical context.

Keywords: behavioural family intervention; depression; objective assessment; therapist competence

Introduction

Ensuring that therapists are competently trained in the delivery of evidence-based therapies is an important first step in delivering effective and efficient psychotherapy to clients. This has been achieved for many individual therapy models such as cognitive behavioural therapy (CBT; McManus *et al.*, 2010; Rakovshik and McManus, 2010; Sharpless and Barber, 2009), but it is not fully achieved for couple and family interventions.

Within the context of depression and relationship distress, there is evidence that behavioural couple therapy (BCT) is effective at eliminating depressive symptoms by targeting general relationship functioning (Baucom *et al.*, 1998; Whisman and Baucom, 2012). Therefore, recent

National Institute for Health and Clinical Excellence (NICE) guidelines for depression have cited BCT as an evidence-based treatment (NICE, 2009).

In England, the Improving Access to Psychological Therapies Programme (IAPT) has recommended additional training for qualified CBT therapists in BCT, in order to increase access to this evidence-based form of therapy (IAPT; Clark *et al.*, 2009). This has included not only BCT where one or both partners are experiencing depression, but also for clinical presentations where there is an interaction between depression, couple functioning and long-term health conditions. Consistent with the move towards competency-based training more broadly (Holloway, 2012), such an expansion in BCT training has also required the creation of a competence framework for couple therapy for depression (Clulow, 2010a), and the need to develop valid and reliable measures of competence in the delivery of BCT.

Reliable and valid measures of therapist competence are essential to assess the level of competence attained during training and to monitor the quality of treatment provision within routine clinical practice (Fairburn and Cooper, 2011; Muse and McManus, 2013). Reliable and valid measures also allow targeted feedback regarding a therapist's strengths and weaknesses, which can be effective in improving competence (McManus *et al.*, 2010; Muse and McManus, 2013). In addition, competence assessment is essential to interpreting the outcomes of effectiveness studies. Finally, there may be a relationship between therapist competence and treatment outcome. Therefore, reliable and valid measures of therapist competence are essential for assessing trainee therapists and for monitoring the quality of treatment provision to ensure that treatment is optimally effective (Muse and McManus, 2013).

What is competence?

Fairburn and Cooper (2011) defined therapist competence as '... the extent to which a therapist has the knowledge and skill required to deliver a treatment to the standard needed for it to achieve its expected effects' (p. 374). Focusing specifically on professional psychologists, Kaslow (2004) discusses eight domains of competence: (a) ethical and legal issues; (b) individual and cultural diversity; (c) scientific foundations and research; (d) assessment; (e) intervention; (f) consultation and interprofessional collaboration; (g) supervision; and (h) professional development. Competence within the intervention domain includes global therapeutic knowledge and skills (i.e. a therapist's ability to independently assess a client's well-being) and limited-domain competence which refers to specific knowledge and skills according to a given therapeutic domain (Barber *et al.*, 2007). This makes it necessary to define model-specific competences.

Importance of measuring therapist competence

Within the field of couple-based intervention, there has historically been a dearth of evaluations of therapist competence (Jacobson and Addis, 1993). Most of the research that does exist in this area has been conducted within CBT, but increasingly attention is being paid to this across the psychotherapies and also in couple-based interventions (Jacobson and Addis, 1993).

The American Association for Marriage and Family Therapy (2004), for example, has identified a set of core competences needed when practising marriage and family therapy. In the UK, research groups have defined a set of specific competences required to deliver effective couple therapy for partners with depression (Clulow, 2010a) as well as a competence framework for systemic therapies (Pilling *et al.*, 2010). Both competence models were based on the CBT competence framework developed by Roth and Pilling (2007) and bring together the competences and skills identified within a range of manuals that are evidence-based and likely to be effective in treating depression (Clulow, 2010b) and other mental health difficulties (Pilling *et al.*, 2010). The frameworks have the same five domains as Roth and Pilling's framework (2007): generic therapeutic competences (e.g. knowledge of depression),

basic competences (e.g. knowledge of sexual functioning in couples, how depression manifests in couples), specific competences (e.g. techniques that engage a couple), specific applications (e.g. BCT) and metacompetences (e.g. using clinical judgement when implementing the therapy).

The need for valid measures

As the first post-qualification CBT course accredited by the BABCP, the BCT training course needed to ensure therapists doing the training and achieving BCT accreditation have reached a certain standard of competence. The most commonly used method of assessing competence within training courses and routine practice is observer ratings of therapists' in-session performance (Muse and McManus, 2013) using a competence rating scale. Therefore, reliable and valid measures of therapist competence are essential when assessing therapist competence (Muse and McManus, 2013). Within couple therapy, competency rating scales are scarce. In an extensive literature search¹, only two competency rating scales for a couple therapy context were found, and neither was appropriate for the purpose for assessing BCT competence specifically. For example, one scale, the 'Couple Therapy for Depression Competency Adherence Scale', was developed by the Tavistock Relationship workgroup who deliver couple therapy training within IAPT (Hewison, 2011). The scale is long, consisting of 41 items. However, it only covers one of the five suggested competence domains (i.e. specific couple therapy techniques; Clulow, 2010a) and includes a comprehensive list of techniques used across all couple therapy for depression approaches (Clulow, 2010b). Not all competencies are expected to be observed during one session. Certain competencies listed relate to different stages of therapy and others are mutually contradictory because they are part of different therapeutic interventions (Hewison, 2011). This would suggest that the assessor needs knowledge of all the different approaches to couple therapy to be able to rate the trainee in a reliable manner. Therefore, using this scale would not be feasible within a BCT-specific training course. Additionally, no psychometric data could be found on how the scale was developed or on the scale's reliability and validity.

The second scale identified was the 'Behavioral Couple Therapy Competence Rating Scale' (Jacobson *et al.*, 2000). This scale was used as part of two larger randomised control trials (Christensen *et al.*, 2004; Jacobson *et al.*, 2000), comparing traditional behavioural couples therapy (TBCT; Jacobson and Margolin, 1979) and integrative behavioural couple therapy (IBCT; Christensen and Jacobson, 1998). This scale was developed for those trials to ensure that the therapist did not display a new treatment bias and hence focused exclusively on TBCT. However, the scale has short anchor points describing the score on the scoresheet, but no accompanying manual and only a specialist consultant observing several of a therapist's sessions would be able to judge if competence is achieved (Jacobson *et al.*, 2000); in addition, the scale did not focus on addressing depression within an interpersonal context. Within routine practice a more practical approach is needed, allowing trainers or supervisors to measure the competence demonstrated in a single therapy session and allowing those other than specialists within the field to apply the scale. Furthermore, the Behavioral Couple Therapy Competence Rating Scale does not cover all aspects of BCT (Epstein and Baucom, 2002), and no psychometric data on the validity or reliability of the scale were found.

BCT is a particular approach to couple therapy for depression requiring specific techniques to be used in a competent manner (Baucom *et al.*, 2015). Therefore, neither of the measures mentioned above could effectively measure therapist competence in the BCT model. BCT

¹An extensive literature review was carried out, using PsychInfo, Science Direct, PsycArticles, Web of Knowledge, MEDLINE and PubMed databases using the broad strategy of including any of the following competence-related terms: 'competency scale', 'therapist competence/y', 'clinical skill', 'assessing competence/y' or 'competency intervention', in combination with the following couples-related terms: 'BCT', 'behavioural couples therapy', 'EBCT', 'enhanced cognitive behavioural therapy for couples', 'CBT for couples', 'cognitive behavioural therapy for couples', 'couples therapy'.

Table 1. BCTS-D compared with the CTS-R

BCTS-D: Domain 1. Structure of the session		Comparable CTS-R items
Item 1	Agenda setting	Agenda setting
Item 2	Review of homework (from the previous session)	Homework setting
Item 14	Setting homework (for the next session) ¹	—
Item 15	Ending the session	Overlap with feedback item
BCTS-D: Domain 2. Interaction with the couple and management of the therapeutic process		
Item 3	Collaboration	Collaboration
Item 4	Facilitating couple communication ²	—
Item 5	Pacing and flow	Pacing and efficient use of time
Item 6	Therapist's interpersonal effectiveness	Interpersonal effectiveness
Item 7	Guided discovery	Guided discovery
Item 13	Dyadic conceptualisation	Conceptual integration
BCTS-D: Domain 3. Interventions selected and employed		
Item 8	Formulation of depression in context ²	—
Item 9	Selection of intervention strategy	Application of change methods
Item 10	Emotion-focused interventions	Eliciting appropriate emotional expression
Item 11	Cognitive interventions	Eliciting key cognitions
Item 12	Behavioural interventions	Eliciting behaviours

¹Homework item split into two items.

²Added to BCTS-D.

requires a unique measure capturing the general skills of couple therapy and the specific skills of BCT, as well as the adherence to the BCT model. As such, developing reliable, valid and usable methods for assessing the competence with which BCT is delivered is crucial to the continued progression of the field. However, future research needs to strike a balance between the need for reliable and valid assessments of therapist competence and the limits on resource availability within routine practice. Cost-effective methods of assessing competence need to be developed further which can be utilised across a range of practice settings (Muse and McManus, 2013).

Scale development

The BCTS-D is based on the established and widely used 'Cognitive Therapy Scale-Revised' (CTS-R; Blackburn *et al.*, 2001). The rationale was twofold. First, it fitted well with the defined couple competences which were based on a CBT competence model (Clulow, 2010a), and, secondly, it was a cost- and time-effective way to use an established measure to define generic competences into which model-specific BCT competences could be integrated.

An expert group consisting of one of the founders of BCT (D.H. Baucom; see Epstein and Baucom, 2002) and several key members of the BCT group in the UK adapted the CTS-R (Blackburn *et al.*, 2001), an observation-based rating scale, to a couple context creating the BCTS-D (Corrie, Fischer, Worrell, & Baucom, n.d.; see Appendix A in Supplementary material). The CTS-R was modified to be appropriate for rating a therapist's degree of competence in BCT working with a couple where at least one partner is experiencing depression. A comprehensive manual providing detailed descriptions of each competence level for every item was also developed. Each item of the CTS-R was rephrased to fit the couple context, two new items ('Focus on Depression in Context' and 'Facilitating Couple Communication') were added, and the homework item was split into two different items (reviewing homework and setting new homework; see Table 1). Thus the scale consists of 15 items. Further, and again consistent with the CTS-R, each item is rated on a 7-point Likert scale and is underpinned by Dreyfus and Dreyfus' (1986) model of competence. The total BCTS-D score is 90 and the overall passing threshold was set at a score of 45 (i.e. 50%, the

same threshold as for the CTS-R). Finally, the BCTS-D was designed to assess both audio and video recordings of active treatment sessions (i.e. excluding assessment and ending sessions).

Study design

In this study, BCT supervisors and BCT trainees were initially asked to rate the competence of an audio recording of a BCT session using the BCTS-D. The study then consisted of two phases. First, formal feedback about the BCTS-D from both BCT supervisors (i.e. experts in the field) and trainees (i.e. relative novices who are likely to receive feedback on the BCTS-D and use the tool to rate their own competence) was collected. This feedback was used to examine content validity (i.e. adequate representation of BCT competence), face validity (i.e. credibility and plausibility of items measuring BCT competence) and perceived usability. Second, a focus group was conducted and qualitative feedback was collected to gain more in-depth feedback on the experience of using the BCTS-D and its usefulness.

Method

Procedure

A workshop for BCT supervisors and trainees was organised where the BCTS-D was introduced and explained during a short presentation. Thereafter, supervisors and trainees were asked to individually rate the competence of a BCT therapist on the BCTS-D by listening to an audio recording of a single therapy session. The therapy session selected for rating was of a mid-treatment session where one partner had been diagnosed with depression. Participants were sent the BCTS-D and the accompanying manual in advance, were asked to familiarize themselves with its content and were advised to use the manual for further guidance when undertaking their ratings. After using the scale, participants filled in a detailed feedback questionnaire exploring the content and face validity of the BCTS-D as well as its usability.

In the last part of the workshop, a focus group was conducted and qualitative data were collected to gain more in-depth feedback on the participants' experiences of using the BCTS-D and its usefulness. An information sheet about the study was included and written consent was obtained.

Participants

The optimal number of participants for analysing the content validity of a rating scale depends on a range of factors such as length and style of the scale and practical considerations including the availability of experts. However, it is generally agreed that using more than five participants facilitates detection and exclusion of rater outliers and increases the robustness of ratings (Haynes *et al.*, 1995).

The study recruited BCT supervisors and trainees (see Table 2 for demographic information) to support the scale development, to gain feedback from the target population and to ensure that it was usable for a variety of stakeholder groups (Brewer and Hunter, 2006; Campanelli *et al.*, 1991). Getting the perspectives of both groups was a way of ensuring that the measure was fit for purpose for future groups of trainees and supervisors for the purposes of formal evaluation, self-assessment and reflection. All participants had either previously completed a BCT training course or were in the process of completing a BCT training course at postgraduate level. Participants were invited to the training day as part of their continuous professional development. An email invitation was sent to 19 supervisors and 44 trainees of which six supervisors (32%) and 14 trainees (32%) attended.

Table 2. Participants' demographics

	Supervisors <i>n</i> = 6	Trainees <i>n</i> = 14
Age [mean (<i>SD</i>)] [range]	44.67 (11.84) [31–61]	38.64 (6.34) [28–50]
CBT accredited	4/6	12/14
Years practising CBT [mean (<i>SD</i>)] [range]	14.5 (6.16) [7–20]	7.0 (2.51) [1–10]
Number of CBT cases		
Treated <50 cases	0	1
Treated 50–200 cases	1	2
Treated >200 cases	5	10
Years practising BCT [mean (<i>SD</i>)] [range]	5.67 (2.34) [2–9]	1.14 (.86) [1–3]
Number of BCT cases		
Treated <50 cases	4	14
Treated 50–200 cases	2	0
Treated >200 cases	0	0
Familiar with CTS-R	4	13
Profession		
Clinical psychologist	4	4
CBT therapist	1	9
Nurse	1	0
Counselling psychologist	0	1

Part 1: Evaluation of content and face validity and perceived usability of the BCTS-D

Material

Content and face validity. Participants were asked to rate the relevance and clarity of each of the items (where '1' was not relevant/clear, '2' was somewhat relevant/clear, '3' was quite relevant/clear and '4' was very relevant/clear). An index of content validity was examined (CVI; Lynn, 1986) by calculating the percentage of participants who rated the item as both relevant and clear (i.e. a rating of 3 or 4 on the 4-point scale). This was done separately for BCT supervisors and trainees as the calculation of the CVI is usually based on expert feedback.

Usability questions. Participants were asked if the scale was useful to judge competence (1: not useful, to 4: very useful) and to rate the overall style, appearance and layout of the scale (1: poor, to 4: very good). They were also asked how easy it was to use the scale (1: not easy, to 4: very easy), if the scale gave the opportunity for useful feedback (1: not useful, to 4: very useful), and how appropriate they found the scoring system (1: not appropriate, to 4: very appropriate). If participants provided a score of 3 or less, they were asked to provide feedback on how they believed that the scale could be improved.

Part 2: Qualitative evaluation of scale utility

Material

The aim of the focus group was to collect in-depth feedback on the strengths and weaknesses of the scale. A semi-structured interview schedule was used to facilitate the discussion. Within the schedule, emphasis was placed on facilitating discussion of both positive and negative responses to the BCTS-D using open-ended questions. Where negative feedback was provided, participants were asked to share their views on whether the issue could be resolved and, if so, how.

The qualitative feedback provided by participants in the questionnaire and focus group was analysed using thematic analysis (Braun and Clarke, 2006). Thematic analysis was chosen because it is flexible and can be used as a descriptive approach reflecting the reality of

participants (Braun and Clarke, 2006). This process consisted of six phases: familiarization with the data, coding generation, searching for patterns based on the initial coding, reviewing themes, defining and naming themes, and producing the report. The process of analysis does not provide in-depth description and interpretation of the data (i.e. no attempt was made to identify the broader meanings and implications of the themes or to relate these themes to previous literature), as the intention is simply to identify ‘surface level’ meaning of the participants’ comments, such as problems with the scale and any suggested improvements (Braun and Clarke, 2006).

Results

Part 1: Evaluation of content and face validity and perceived usability of the BCTS-D

Content and face validity

Content validity scores are presented in Table 3 for both BCT supervisors and trainees, who found all items in the scale relevant and clear (a rating of 3 or 4 on the 4-point scale), with no significant differences between the scores assigned by supervisors and trainees. The CVI (Lynn, 1986) was above the suggested threshold of 70% for all items. Only one supervisor out of six and none of the trainees indicated that there were important aspects of BCT competence they felt were missing from the scale. Mann–Whitney tests² revealed no significant differences between the scores for relevance or clarity assigned by trainees and supervisors, indicating that trainees and supervisors did not differ significantly in their views. Thus, all items in the scale were viewed as having acceptable content validity by both supervisors and trainees.

Usability

All participants rated the scale as at least ‘quite useful’ for judging BCT competence. The way the scale enables the provision of feedback was also rated at least as ‘quite’ useful. In addition, the results indicate the scale was at least ‘quite’ easy to use, had at least ‘good’ style, appearance and layout, and had at least a ‘quite’ appropriate scoring system with no differences between trainees and supervisors. Mann–Whitney tests revealed no significant differences between the scores for relevance or clarity assigned by trainees and supervisors, indicating that trainees and supervisors did not differ significantly in their views. Thus, all items in the scale were viewed as having acceptable content validity by both supervisors and trainees (see Table 4).

Part 2: Qualitative evaluation of scale utility

Thematic analysis identified major themes in the qualitative feedback on how to improve the scale. Additionally, the majority of participants commented positively on the development of the scale. Specifically, participants believed that the scale would aid self-reflection and personal development and could be used as a self-directed learning tool to facilitate a more objective review of one’s ability, while being reminded of the core competences that enable fidelity to the BCT model. Furthermore, it was stated that the scale is very thorough and gives the opportunity to rate a therapist’s competence while accommodating the couple’s needs (i.e. being highly directive if needed). However, these comments were not included in Table 5 as they did not add anything to how the scale could be improved further.

From the coding and the thematic analysis of the written feedback and the focus group data, four main themes emerged: (a) need to capture competence better; (b) complexity of competence assessment ratings; (c) improve clarity on how to use the scale (i.e. revisions needed to make the

²As the data were not normally distributed, non-parametric Mann–Whitney *U*-tests were performed to examine the difference between the scores assigned by trainees and supervisors.

Table 3. Content validity results for each item in the BCTS-D

	CVI ¹		Relevance (1–4)		Mann–Whitney U^2	Clarity (1–4)		Mann–Whitney U^2
	Trainees (<i>n</i> = 14)	Supervisors (<i>n</i> = 6)	Trainees (<i>n</i> = 14)	Supervisors (<i>n</i> = 6)		Trainees (<i>n</i> = 14)	Supervisors (<i>n</i> = 6)	
	(%)	(%)	Mean (<i>SD</i>)	Mean (<i>SD</i>)		Mean (<i>SD</i>)	Mean (<i>SD</i>)	
1 (agenda setting)	92.86	83.33	3.57 (.76)	3.67 (.82)	$p = .779$	3.64 (.5)	3.67 (.82)	$p = .659$
2 (review of homework)	100	100	3.86 (.43)	4.0 (.0)	$p = .659$	3.79 (.43)	3.67 (.52)	$p = .718$
3 (collaboration)	100	100	3.93 (.27)	3.83 (.41)	$p = .779$	3.71 (.47)	3.67 (.52)	$p = .904$
4 (facilitating couple communication)	100	91.67	3.93 (.27)	4.0 (.0)	$p = .841$	3.64 (.5)	3.33 (.82)	$p = .494$
5 (pacing and flow)	100	100	3.71 (.47)	3.83 (.41)	$p = .718$	3.64 (.5)	3.83 (.41)	$p = .547$
6 (therapist's interpersonal effectiveness)	100	100	3.93 (.27)	4.0 (.0)	$p = .841$	3.78 (.43)	3.83 (.41)	$p = .904$
7 (guided discovery)	100	91.67	3.85 (.38)	3.83 (.41)	$p = .999$	3.69 (.48)	3.17 (1.17)	$p = .416$
8 (formulation of depression in context)	96.43	100	3.77 (.44)	4.0 (.0)	$p = .467$	3.46 (.66)	3.67 (.52)	$p = .639$
9 (selection of intervention strategy)	96.43	100	3.92 (.28)	4.0 (.0)	$p = .831$	3.62 (.65)	3.83 (.41)	$p = .639$
10 (emotion-focused interventions)	100	100	3.92 (.29)	4.0 (.0)	$p = .820$	3.33 (.49)	3.83 (.41)	$p = .102$
11 (cognitive interventions)	96.43	100	3.79 (.43)	4.0 (.0)	$p = .494$	3.36 (.63)	3.83 (.41)	$p = .153$
12 (behavioural interventions)	100	100	3.93 (.27)	4.0 (.0)	$p = .841$	3.5 (.52)	3.83 (.41)	$p = .274$
13 (dyadic conceptualisation)	96.15	91.67	3.77 (.44)	4.0 (.0)	$p = .467$	3.54 (.88)	3.5 (.84)	$p = .898$
14 (setting homework)	100	100	3.93 (.27)	4.0 (.0)	$p = .841$	3.86 (.36)	4.0 (.0)	$p = .659$
15 (ending the session)	100	100	3.92 (.28)	4.0 (.0)	$p = .831$	3.85 (.38)	4.0 (.0)	$p = .639$

¹CVI: content validity index, the percentage of participants who rated item as 3 or 4 on the 4-point scale (1: *not* to 4: *very*) for **both** relevance and clarity.²As the data were not normally distributed, non-parametric Mann–Whitney U -tests were performed to examine the difference between the scores assigned by trainees and supervisors.

Table 4. Usability results for the BCTS-D

	Trainees <i>n</i> = 14	Supervisors <i>n</i> = 6	Mann-Whitney <i>U</i>
Useful to judge competence (1–4) [mean (<i>SD</i>)]	3.5 (.52)	3.83 (.41)	<i>p</i> = .274
Opportunity for useful feedback (1–4) [mean (<i>SD</i>)]	3.29 (.47)	4.0 (0)	<i>p</i> = .051
Style, appearance and layout (1–4) [mean (<i>SD</i>)]	3.5 (.52)	3.83 (.41)	<i>p</i> = .274
Appropriate scoring system (1–4) [mean (<i>SD</i>)]	3.43 (.51)	3.8 (.45)	<i>p</i> = .257
Ease of use (1–4) [mean (<i>SD</i>)]	3.36 (.5)	3.5 (.84)	<i>p</i> = .494

rating of the scale easier); and (d) overlap of items (i.e. aspects of competence assessed in one item overlaps with the aspects of competence assessed in another item). Examples of comments are provided in Table 5, with the full thematic analysis and codes provided in Appendix B in Supplementary material.

The revisions to the BCTS-D made in response to each coding, or an explanation as to why the issue in the theme was not resolved, are also outlined in Table 5.

Discussion

The results of these two studies suggest that the BCTS-D may have good validity and usability. The BCTS-D received encouraging feedback from the research participants. The majority of BCT supervisors and trainees found all items on the scale relevant and clear, and only a very small percentage of participants indicated that items in the scale inappropriately overlapped with other items. Only one supervisor of the 20 participants indicated that important aspects of BCT competence were missing from the scale. Overall, supervisors and trainees indicated that they found the scale useful and easy to use with an appropriate scoring system. The qualitative feedback collected reflected a similar picture with many positive comments such as the scale would be useful for self-reflection, personal development, and can be used to aid an objective review of one's competences.

Some feedback indicated some confusion regarding how to use the BCTS-D and its manual. At this point, it is uncertain whether this is due to the manual being insufficiently clear or attributable to participants not having had sufficient opportunity to familiarise themselves with it prior to participating in the study. These findings may, however, be representative in that those using the BCTS-D in research, training and routine practice settings might find it unduly arduous to read the manual, which consists of 98 pages. Consequently, participants identified the opportunity to practise and become familiar with the scale before using it as important. Ideally, a training day before the feedback workshop would have taken place to ensure that all participants had had time to read the manual. However, due to financial and time constraints, this proved to be unrealistic. Another explanation for certain areas of confusion could be that BCT is a principle-driven approach using more formulation-based techniques, and many of the participants had only recently started their BCT training after working in IAPT services which are typically organised around delivering treatment protocols.

A further recommendation was to add a contextual framework to the BCTS-D. The failure to take therapeutic context into account when measuring competence has been criticised by Waltz and colleagues (1993), who stated that it undermines the validity of any scale. This could lead to therapists being penalized and rated as less competent for treating more complex cases (Rakovshik and McManus, 2010). To counteract this, use of the BCTS-D requires that therapists submit a summary sheet including demographic information such as each partner's age, a description of the couple's presenting problems, and the aims of the session (see Appendix C in Supplementary material). This information enables the rater to evaluate the session in its broader context.

Table 5. Examples of qualitative feedback on the BCTS-D

Themes	Codes	Response or solution
Complexity of competence assessment ratings	Assess the essence of the session and overall mark for pass/fail; this could be rated by both supervisors and trainees	Factors that conflate people's ratings were highlighted by the expert group such as the rater's overall feeling about a session. Adding an item about the overall session would therefore represent a possible bias (I like the therapist, the session felt comfortable). Furthermore, the scale is designed to define the competence during that session only as competence might fluctuate over time and situation (Dobson and Singer, 2005)
	Complex and long manual which carries the risk that people will not use it	The BCTS-D manual consists of 98 pages as the scale includes 15 items. Assessing competence is very complex and therefore asks for a lot of guidance, which is why the manual is long. Furthermore, the expert group highlighted that training on how to use the scale plus support during supervision will be essential for its correct use and to overcome the barrier of not reading the manual
Complexity of competence assessment ratings	Some uncertainty on how to rate some items as more use of formulation-based adaptations during therapy sessions	BCT is a principle-driven approach rather than a manualised intervention. In order to prove optimally clinically responsive to the needs of a couple, therefore, therapists need to be able to adapt the delivery of the session. Such adaptations are deemed to be illustrative of good practice (e.g. slowing the pace of a session or being more direct when a couple is arguing). Therefore, the manual purposely was not too descriptive on how to rate certain items and ensures that therapists are not marked down for adjusting their approach when needed
	How much training needed before able to do assessment rating with BCTS-D	The expert group are in the process of creating a standardised training session. However, they also emphasise that regular support during supervision on how to use the scale will be important in order to be able to use the scale reliably
Improve clarity of how to rate certain items	Clarification on scoring needed – how to rate items that are not present for appropriate reasons (i.e. items 10, 11, 12: emotion-focused, cognitive and behavioural interventions)	The BCTS-D manual explains that if an item is omitted due to appropriate reasons, it is to be given a pass (scored a 3). The expert group discussed that it might be useful to highlight this scoring rule directly on the scoresheet
	Asking to add more examples on how to score certain items	The research group does not want to make the manual longer, and does not want to add examples to the scoring sheet as it increases the risk of people not using the manual while scoring someone on the BCTS-D
Overlap of items	Unclear purpose of item 7 (guided discovery) as all aspects covered in other items 3, 6 and 11 as well (collaboration, interpersonal effectiveness and cognitive intervention)	The expert group considered this issue and decided to analyse inter-rater reliability to see if item 7 is rated reliably between assessors or not before possibly omitting it from the scale
	Consider cultural influences as its own item or part of interpersonal effectiveness, rather than being part of item 3 (collaboration)	The expert group discussed how cultural influences would normally be considered in item 6 (interpersonal effectiveness), but not as a separate item. The expert group is considering rewording item 6 to make this more obvious in the manual

Despite some issues that will need considering in future research, the scale was well-liked by the participants in the group, who thought it a beneficial development for learning and self-reflection, especially within the BCT training course.

Future development of the BCTS-D

The BCTS-D will require further evaluation as the sample size for this pilot analysis of validity and utility was small. Independent use and verification of the scale will be important as the participant sample was part of the BCT postgraduate diploma programme that has developed the BCTS-D. It is also worth noting that the qualitative feedback collected in this study might have been influenced by demand characteristics. Any such influences would be more likely to have influenced trainees, who were less experienced and less senior than the supervisors. Nonetheless, no differences in the questionnaire ratings assigned by trainees and supervisors were found. Most ratings were very favourable and some aspects of the scale were rated a 4 (i.e. very relevant and very clear) by all six supervisors. This raises the question of whether some of the feedback was positively skewed due to the participant sample. All participants invited to the workshop were from the BCTS-D training group. Those attending were more interested in the development of the BCTS-D and possibly viewed it in a favourable light. However, there is one non-significant trend ($p = .051$) where trainee participants rated 'the opportunity for useful feedback' less favourably than the supervisors. The trainees' average rating was still at 'quite useful' (3.29) and no negative qualitative comments regarding this point were identified in the analysis to help explain the phenomenon. Furthermore, steps were taken to mitigate against participants providing positively biased feedback (i.e. anonymised questionnaires and inviting negative as well as positive comments to help develop the scale further).

Scale development is an iterative process with data collected in later stages being used to improve certain previous steps (DeVellis, 2012). Therefore, the future modification of the scale should be considered and more qualitative feedback collected and quantitative research conducted. Conducting a factor analysis would solve the question of how many constructs underlie the BCTS-D. In addition to the factor analysis, future research should focus on using several recordings, rated by at least two raters to explore the inter-rater reliability of the scale. Unfortunately, it was not possible within the time frame of this study to address the stringent data protection requirements of the National Health Service. Hence therapy recordings were not allowed to be kept in order for competence ratings to be carried out at a later date by an independent rater. Moreover, it was beyond the scope of this study to determine discriminant validity of the BCTS-D. It is expected that therapeutic competence will increase over a year-long course as trainee therapists develop their skills (McManus *et al.*, 2010; Williams *et al.*, 1991). Measuring discriminant validity would indicate if the BCTS-D can provide a useful tool for measuring therapists' progress within a BCT training programme, and this analysis should be included in future research.

Research suggests that familiarisation through reading a manual is not enough to ensure good consistency between raters (Gordon, 2007; Reichelt *et al.*, 2003). Thorough training is proposed as key to ensuring inter-rater reliability. A possibility would be to offer some online training to help familiarisation. The development of a standardized training in how to use the BCTS-D and evaluating its effect on the reliability and perceived validity and usability of the scale, is also essential. Ideally, the training should be standardised to increase inter-reliability including regular discussion and a feedback session about how to calibrate ratings of certain treatment sessions.

Following the qualitative feedback, a summary sheet which includes demographic information about the couple, their presenting problems, diagnosis, formulation, number of sessions, what was done previously and the aim of the session, will be added to the official BCTS-D pack as therapeutic context is an important factor when measuring competence (Stiles *et al.*, 1998;

Waltz *et al.*, 1993). Moreover, the manner in which this contextual information should influence ratings will need to be detailed in the manual.

Finally, to be able to use the scale to measure competence within the training for which it was designed (i.e. a postgraduate diploma in BCT), it will be important to establish an empirically informed cut-off point to determine when competence has been achieved. This will need to take the purpose of the assessment into account as, for example, the threshold to pass an introductory BCT training programme may be lower than the requirement at the end of a one-year BCT training course (Muse and McManus, 2013). Future research could usefully examine whether the BCTS-D can discriminate between BCT trainees and experienced BCT therapists as a first step towards validating that cut-off point.

Conclusion

To ensure good quality treatment within routine practice, valid and reliable competence assessment tools are essential (Muse and McManus, 2013). The newly developed BCTS-D is filling the gap within BCT. This initial evaluation of the BCTS-D demonstrates good face validity, content validity and usability, and suggests that it is a tool useful for providing formative feedback and promoting self-reflection. Hence the BCTS-D appears to have good potential to be a suitable competence assessment scale within clinical practice, training settings and research studies. Several issues will need to be addressed in future research to optimise the use of the scale within a BCT training context. Furthermore, it will be essential to analyse the psychometric properties of the BCTS-D within a real-world clinical setting in order to confirm its validity and the reliable use of the scale. However, despite its limitations, this tool provides an important, initial contribution to helping supervisors and trainees better understand, conceptualise and evaluate competence in this emerging specialism.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/S1754470X20000276>

Acknowledgements. The authors sincerely acknowledge the team of authors (I.-M. Blackburn, I. A. James, D. L. Milne, C. Baker, S. Standart, A. Garland and F. K. Reichelt) who produced the Cognitive Therapy Scale-Revised (2001) upon which the BCTS-D was based. They kindly gave their permission to adapt the CTS-R.

Conflicts of interest. I. Rudolf von Rohr, S. Corrie, M. S. Fischer, D. H. Baucom, M. Worrell and H. Pote have no conflicts of interest with respect to this publication.

Financial support. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Ethical statements. The authors have abided by the Ethical Principles of Psychologists and Code of Conduct as set out by the APA. Ethical permission to collect data within the Central London Training Centre was obtained from the Health Research Authority (HRA, IRAS project ID: 199914) and the Royal Holloway University of London Ethics Committee.

Key practice points

- (1) To ensure good quality treatment within routine practice, valid and reliable competence assessment tools are essential (Muse and McManus, 2013). The newly developed BCTS-D is filling the gap within BCT.
- (2) This initial evaluation of the BCTS-D demonstrates good face validity, content validity and usability, and suggests that it is a tool useful for providing formative feedback and promoting self-reflection. Hence the BCTS-D appears to have good potential to be a suitable competence assessment scale within clinical practice, training settings and research studies.
- (3) Furthermore, it will be essential to analyse the psychometric properties of the BCTS-D within a real-world clinical setting in order to confirm its validity and the reliable use of the scale.
- (4) However, despite its limitations, this research makes an important initial contribution to helping supervisors and trainees better understand, conceptualise and evaluate competence in this emerging specialism.

Further reading

- Baucom, D. H., Fischer, M. S., Corrie, S., Worrell, M., & Boeding, S. E.** (2020). *Treating Relationship Distress and Psychopathology in Couples: A Cognitive-Behavioural Approach*. Abingdon, UK: Routledge.
- Baucom, D. H., Fischer, M. S., Worrell, M., Corrie, S., Belus, J. M., Molyva, E., & Boeding, S. E.** (2018). Couple-based intervention for depression: an effectiveness study in the National Health Service in England. *Family Process*.
- Muse, K., & McManus, F.** (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review*, 33, 484–499. <https://doi.org/10.1016/j.cpr.2013.01.010>.

References

- American Association for Marriage and Family Therapy** (2004). Marriage and family therapy core competencies. Retrieved from: [https://www.coamfte.org/Documents/COAMFTE/Accreditation%20Resources/MFT%20Core%20Competencies%20\(December%202004\).pdf](https://www.coamfte.org/Documents/COAMFTE/Accreditation%20Resources/MFT%20Core%20Competencies%20(December%202004).pdf)
- 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. <https://doi.org/10.1037/0735-7028.38.5.493>.
- Baucom, D. H., Epstein, N. B., Kirby, J. S., & LaTaillade, J. J.** (2015). Cognitive-behavioral couple therapy. In A. S. Gurman, J. Lebow, & D. K. Snyder (eds), *Clinical Handbook of Couple Therapy* (5th edn, pp. 23–60). New York, USA: Guilford Press.
- Baucom, D. H., Shoham, V., Mueser, K. T., Daiuto, A. D., & Stickle, T. R.** (1998). Empirically supported couple and family interventions for marital distress and adult mental health problems. *Journal of Consulting and Clinical Psychology*, 66, 53. <https://doi.org/10.1037/0022-006X.66.1.53>.
- Blackburn, I.-M., James, I. A., Milne, D. L., Baker, C., Standart, S., Garland, A., & Reichelt, F. K.** (2001). The revised cognitive therapy scale (CTS-R): psychometric properties. *Behavioural and Cognitive Psychotherapy*, 29, 431–446. <https://doi.org/10.1017/S1352465801004040>.
- Braun, V., & Clarke, V.** (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77–101. <https://doi.org/10.1191/1478088706qp0630a>.
- Brewer, J., & Hunter, A.** (2006). *Foundations of Multimethod Research: Synthesizing Styles*. Thousand Oaks, CA, USA: Sage Publications.
- Campanelli, P. C., Martin, E. A., & Rothgeb, J. M.** (1991). The use of respondent and interviewer debriefing studies as a way to study response error in survey data. *The Statistician*, 40, 253–264. <https://doi.org/10.2307/2348278>.
- Clark, D. M., Layard, R., Smithies, R., Richards, D. A., Suckling, R., & Wright, B.** (2009). Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. *Behaviour Research and Therapy*, 47, 910–920. <https://doi.org/10.1016/j.brat.2009.07.010>.
- Clulow, C.** (2010a). *Couple Therapy for Depression Competency Framework*. Retrieved from: www.iapt.nhs.uk
- Clulow, C.** (2010b). *The competences required to deliver effective couple therapy for partners with depression*. Retrieved from: https://www.ucl.ac.uk/pals/research/cehp/research-groups/core/pdfs/Couples_Therapy_For_Depression/Couples_Therapy_clinician_guide.pdf
- Christensen, A., Atkins, D. C., Berns, S., Wheeler, J., Baucom, D. H., & Simpson, L. E.** (2004). Traditional versus integrative behavioral couple therapy for significantly and chronically distressed married couples. *Journal of Consulting and Clinical Psychology*, 72, 176–191. <https://doi.org/10.1037/0022-006X.72.2.176>.
- Christensen, A., & Jacobson, N. S.** (1998). *Acceptance and Change in Couple Therapy: A Therapist's Guide to Transforming Relationships*. New York, USA: Norton.
- Corrie, S., Fischer, M. S., Worrell, M., & Baucom, D. H.** (n.d.). Behavioural Couple Therapy Scale – Depression (BCTS-D): The Manual. [In preparation]
- DeVellis, R. F.** (2012). *Scale Development: Theory and Applications* (Vol. 26). Newbury Park, CA, USA: Sage Publications.
- Dobson, K. S., & Singer, A. R.** (2005). Definitional and practical issues in the assessment of treatment integrity. *Clinical Psychology: Science and Practice*, 12, 384–387.
- Dreyfus, H., & Dreyfus, S.** (1986). *Mind over Machine: The Power of Human Intuition and Expertise in the Era of Computer*. New York, USA: Free Press.
- Epstein, N. B., & Baucom, D. H.** (2002). Enhanced cognitive-behavioral therapy for couples: a contextual approach. American Psychological Association. Retrieved from: <http://psycnet.apa.org/psycinfo/2003-04003-000>
- Fairburn, C. G., & Cooper, Z.** (2011). Therapist competence, therapy quality, and therapist training. *Behaviour Research and Therapy*, 49, 373–378. <https://doi.org/10.1016/j.brat.2011.03.005>.
- Gordon, P. K.** (2007). A comparison of two versions of the Cognitive Therapy Scale. *Behavioural and Cognitive Psychotherapy*, 35, 343–353. <https://doi.org/10.1017/S1352465806003390>.
- Haynes, S. N., Richard, D., & Kubany, E. S.** (1995). Content validity in psychological assessment: a functional approach to concepts and methods. *Psychological Assessment*, 7, 238. <https://doi.org/10.1037/1040-3590.7.3.238>.
- Hewison, D.** (2011). Curriculum for Couple Therapy for Depression. National IAPT Programme Team.

- Holloway, E. L.** (2012). Professional competence in supervision. In J. N. Fuertes, A. A. Spokane, & E. L. Holloway (eds). *Speciality Competencies in Counseling Psychology* (pp. 165–181). New York, USA: Oxford University Press.
- Jacobson, N. S., & Addis, M. E.** (1993). Research on couples and couple therapy: what do we know? Where are we going? *Journal of Consulting and Clinical Psychology*, *61*, 85–93. <https://doi.org/10.1037/0022-006X.61.1.85>.
- Jacobson, N. S., Christensen, A., Prince, S. E., Cordova, J., & Eldridge, K.** (2000). Integrative behavioral couple therapy: an acceptance-based, promising new treatment for couple discord. *Journal of Consulting and Clinical Psychology*, *68*, 351–355. <https://doi.org/10.1037/0022-006X.68.2.351>.
- Jacobson, N. S., & Margolin, G.** (1979). *Marital Therapy: Strategies Based on Social Learning and Behavior Exchange Principles*. Psychology Press.
- Kaslow, N. J.** (2004). Competencies in professional psychology. *American Psychologist*, *59*, 774. <https://doi.org/10.1037/0003-066X.59.8.774>.
- Lynn, M. R.** (1986). Determination and quantification of content validity. *Nursing Research*, *35*, 382–386. <https://doi.org/10.1097/00006199-198611000-00017>.
- 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. <https://doi.org/10.1016/j.brat.2010.08.002>.
- Muse, K., & McManus, F.** (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review*, *33*, 484–499. <https://doi.org/10.1016/j.cpr.2013.01.010>.
- National Institute for Clinical Excellence (NICE)** (2009). Depression in adults: recognition and management (Clinical Guideline 90). London, UK: NICE.
- Pilling, S., Roth, A., & Stratton, P.** (2010). The competences required to deliver effective systemic therapies. Retrieved from: https://www.researchgate.net/publication/260320937_The_competences_required_to_deliver_effective_Systemic_Therapies
- Rakovshik, S. G., & McManus, F.** (2010). Establishing evidence-based training in cognitive behavioral therapy: a review of current empirical findings and theoretical guidance. *Clinical Psychology Review*, *30*, 496–516. <https://doi.org/10.1016/j.cpr.2010.03.004>.
- Reichelt, F. K., James, I. A., & Blackburn, I.-M.** (2003). Impact of training on rating competence in cognitive therapy. *Journal of Behavior Therapy and Experimental Psychiatry*, *34*, 87–99. [https://doi.org/10.1016/S0005-7916\(03\)00022-3](https://doi.org/10.1016/S0005-7916(03)00022-3).
- Roth, A. D., & Pilling, S.** (2007). The competences required to deliver effective cognitive and behavioural therapy for people with depression and with anxiety disorders. Retrieved from: <http://discovery.ucl.ac.uk/137911/>
- Sharpless, B. A., & Barber, J. P.** (2009). A conceptual and empirical review of the meaning, measurement, development, and teaching of intervention competence in clinical psychology. *Clinical Psychology Review*, *29*, 47–56. <https://doi.org/10.1016/j.cpr.2008.09.008>.
- Stiles, W. B., Honos-Webb, L., & Surko, M.** (1998). Responsiveness in psychotherapy. *Clinical Psychology: Science and Practice*, *5*, 439–458.
- Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S.** (1993). Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, *61*, 620. <https://doi.org/10.1037/0022-006X.61.4.620>.
- Whisman, M. A., & Baucom, D. H.** (2012). Intimate relationships and psychopathology. *Clinical Child and Family Psychology Review*, *15*, 4–13. <https://doi.org/10.1007/s10567-011-0107-2>.
- Williams, R., Moorey, S., & Cobb, J.** (1991). Training in cognitive-behaviour therapy: pilot evaluation of a training course using the Cognitive Therapy Scale. *Behavioural Psychotherapy*, *19*, 373–376. <https://doi.org/10.1017/S0141347300014075>.